Dissertation zur Erlangung des Doktorgrades
der Fakultät Chemie und Pharmazie
der Ludwig–Maximilians–Universität München

Development of Low-Scaling Methods to Calculate Ground State Energies and Analytical Gradients Based on the Adiabatic-Connection Fluctuation-Dissipation Theorem

Matthias Georg Beuerle

aus

Bamberg

2019

Erklärung
Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Christian Ochsenfeld betreut.

Eidesstattliche Versicherung
Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 27.06.2019 (Matthias Beuerle)

Dissertation eingereicht am: 16.04.2019

1. Gutachter: Prof. Dr. Christian Ochsenfeld

2. Gutachter: Prof. Dr. Regina de Vivie-Riedle

Mündliche Prüfung am: 18.06.2019

Danksagung

Zuallererst bedanke ich mich bei Prof. Dr. Christian Ochsenfeld für die Möglichkeit meine Doktorarbeit in seinem Arbeitskreis anfertigen zu können, die spannende Themenstellung und die Unterstützung im Laufe der Jahre.

Desweiteren bedanke ich mich bei Prof. Dr. Regina de Vivie-Riedle für die Anfertigung des Zweitgutachtens.

Im AK Ochsenfeld bedanke ich mich für die produktive Zusammenarbeit insbesondere bei Dr. Henry Schurkus, Dr. Arne Luenser, Daniel Graf und Dr. Jörg Kussmann. Weiterhin bedanke ich mich für anregende Diskussionen fachlicher und nicht fachlicher Natur, sowie sportliche und soziale Aktivitäten bei Dr. Sigurd Vogler, Travis Thompson, Henryk Laqua und dem Rest der Arbeitsgruppe.

Bei meiner Familie und meiner Freundin Katja bedanke ich mich für die ständige Unterstützung während der Promotion und meines sonstigen Lebens.

Summary

New methods to efficiently calculate energetics and first order-properties for mean-field and correlated electronic structure theories are presented. In linear-scaling short-range hybrid calculations new integral screening criteria exploiting the locality of the attenuated Coulomb operator are introduced that allow to significantly increase the performance of these density functionals. This enables short-range hybrid calculations at a similar cost as pure semi-local density functional theory (DFT) calculations with increased accuracy due to the admixture of exchange. At the level of correlated electronic structure theory, the realm of systems which can be calculated with a linear-scaling random-phase-approximation (RPA) method is extended with a novel multi-node parallel algorithm. Furthermore, more efficient quadrature schemes are used based on the equivalence of the employed integral transforms with the Fourier transform of the non-interacting polarizability. In combination with a new way to introduce Cholesky orbitals this results in a more effient and numerically more accurate linear-scaling RPA correlation energy method with improved memory requirements. Combining these techniques further with an approach to calculate analytical first order properties using quantities from many body perturbation theory, allows to present a low-scaling method that gives access to RPA gradients for molecular systems with several hundred atoms. Moreover linear- and low-scaling methods to calculate different beyond RPA correlation energies are devised, exploiting the locality of exchange type contractions. This enables to apply these more accurate RPA methods to significantly larger systems, which allows to demonstrate the gain in accuracy for large, dispersion dominated systems. The newly developed methods extend the scope of RPA and RPA with exchange correlation energies, and first order RPA properties significantly, while the accuracy is under full numerical control. Furthermore, there is no overhead to the respective canonical algorithm, making the presented methods competitive also for small system sizes. Finally, a new beyond RPA scheme is presented that combines the benefits of plain RPA and RPA with second-order screened exchange, leading to a more balanced, highly accurate post Kohn-Sham method.

List of Publications

This is a cumulative dissertation, comprising five articles in peer-reviewed journals (I-V). In the following, all articles are stated together with the author's contribution to each of them.

I **M. Beuerle**, J. Kussmann, C. Ochsenfeld,
"Screening Methods for Linear-Scaling Short-Range Hybrid Calculations on CPU and GPU Architectures",
J. Chem. Phys., **146**, 144108 (2017).
Contribution by the Author: *Most of the derivation, most of the implementation, all calculations and writing the manuscript*

II D. Graf, **M. Beuerle**, H. F. Schurkus, A. Luenser, G. Savasci, C. Ochsenfeld,
"Accurate and Efficient Parallel Implementation of an Effective Linear-Scaling Direct Random Phase Approximation Method",
J. Chem. Theory Comput., **14**, 2505 (2018).
Contribution by the Author: *Parts of the derivation, assistance with the implementation and assistance with writing the manuscript.*

III **M. Beuerle**, C. Ochsenfeld,
"Low-Scaling Analytical Gradients for the Direct Random Phase Approximation using an Atomic Orbital Formalism",
J. Chem. Phys., **149**, 244111 (2018).
Contribution by the Author: *All of the derivation, all of the implementation, all calculations and writing the manuscript.*

IV **M. Beuerle**, C. Ochsenfeld,
"Short-Range Second Order Screened Exchange Correction to RPA Correlation Energies",
J. Chem. Phys., **147**, 204107 (2017).
Contribution by the Author: *All of the derivation, all of the implementation, all calculations and writing the manuscript.*

V **M. Beuerle**, D. Graf, H. F. Schurkus, C. Ochsenfeld,
"Efficient Calculation of Beyond RPA Correlation Energies in the Dielectric Matrix Formalism",
J. Chem. Phys., **148**, 204104 (2018).
Contribution by the Author: *All of the derivation, most of the implementation, all calculations and most of writing the manuscript.*

Publications not related to this thesis:

VI **M. Beuerle**, Neil P. Dufton, Anna M. Randi and Ian R. Gould
"Molecular dynamics studies on the DNA-binding process of ERG"
Mol. BioSyst., **12**, 3600, (2016).

Contents

Chapter 1

Introduction

The overarching goal of quantum chemistry is to reliably predict or complement experimental data. The two requirements that result from this goal are accurate theoretical methods and the availability of efficient algorithms to obtain valuable insights for large and complex systems. The perhaps most central equation of quantum chemistry is the Schrödinger equation [1] in the Born-Oppenheimer approximation [2]. While a route to the numerically exact solution of the Schrödinger equation has been known for decades by the means of Full-CI (FCI), this approach is limited to the smallest molecules and atoms due to its prohibitive computational cost. On the other side of the spectrum, the Hartree-Fock (HF) method [3–5], as the simplest approach to an approximate solution of the Schrödinger equation, is nowadays routinely applicable to systems with thousands of atoms. The HF method fails, however, to accurately describe experimental observables.

Quantum chemistry has been focussing on bridging the gap between accuracy and computational efficiency. Two main avenues have emerged to tackle this problem. On the one hand, wave function based methods aim for a correction to the HF Slater determinant or use a more sophisticated ansatz for the wave function. As they are derived directly from first principles and do not contain any empirical parameters, they are typically highly accurate and can be benchmarked reliably. The downside of these methods is that they are usually computationally very involved. On the other hand, density functional theory (DFT) in its simplest form has a comparable computational cost to HF. While there are areas of applications where the accuracy of DFT rivals the accuracy of wave function based methods, DFT has a variety of limitations. As DFT functionals often contain fitted parameters, their transferability can be limited [6]. Furthermore, there are intrinsic problems that seem to be uncircumventable in the realm of pure DFT such as the lacking description of dispersion [7, 8], or in general non-local phenomena.

In the past years, methods based on the adiabatic-connection fluctuation-dissipation theorem (ACFDT) [9–11], the simplest being the random-phase-approximation (RPA) [12], have regained popularity as electronic-structure methods. While the adiabatic-connection (AC) is a concept which has been used widely to derive DFT functionals, these methods also show characteristics of wave function based methods. The ACFDT methods are derived from first principles and contain no empirical parameters. In comparison to other wave function methods the RPA in its modern

formulation [13] is a fairly low-cost electron correlation method improving certain problems of DFT, e.g., the abovementioned limited transferability and the problems in describing dispersion effects [14, 15]. Furthermore, using tools from time-dependent DFT or many-body perturbation theory, more sophisticated ACFDT methods can be derived that improve upon the RPA. The first step is the inclusion of exchange effects. This makes ACFDT methods an interesting field at the border of DFT and wave function theory giving access to a class of non-empirical, highly accurate DFT functionals with reasonable computational cost.

Before the present work, a formulation of RPA correlation energies [16, 17] has been introduced where the computation time scales linearly with the system size, allowing for the treatment of systems with up to 1000 atoms. More sophisticated ACFDT methods, such as RPA with exchange schemes and properties at the RPA level of theory were still limited to small systems. This thesis extends the field of ACFDT methods in three directions. First, it improves the numerical accuracy and further extends the range of systems accessible to RPA correlation energies. Second, it presents an efficient method to calculate analytical first order properties at the RPA level of theory for molecules with hundreds of atoms, which is indispensable for theoretical studies. Third, it extends the class of RPA with exchange methods by introducing a new RPA with short-range second-order screened exchange functional, and presents linear and low-scaling methods to calculate beyond RPA correlation energies for large systems. Finally, not in the field of ACFDT, new screening methods are devised allowing for more efficient, linear-scaling short-range hybrid calculations, making this class of DFT functionals a more accurate, yet computationally similarly costly alternative to pure DFT.

This work is a cumulative dissertation with **Publications I-V** in Chapter 3 being the main part of this disseration. In Chapter 2, an overview of modern electronic structure theory including the theoretical foundations of this work is given. Furthermore, the techniques employed to obtain the efficent, yet accurate electronic structure methods presented in this work are introduced. Finally, Chapter 4 concludes this work. A short introduction to each publication is given below.

In **Publication I** new screening methods are introduced to significantly optimize the efficiency of short-range hybrid DFT calculations, a class of DFT functionals that employ a fraction of short-range exact-exchange. The calculation of the short-range exact-exchange contribution increases the accuracy in comparison to pure DFT functionals. The calculation of short-range exact-exchange is then, however, also the most expensive step of the calculation. Combining new integral estimates for the arising short-range electron repulsion integrals with state-of-the-art linear-scaling exchange methods significantly reduces the computational demand of short-range hybrid DFT, making it similar in cost to the less accurate pure DFT methods. Screening schemes for both classical CPU and GPU computing architectures are presented.

The remainder of this thesis is concerned with developments in the field of ACFDT methods. **Publication II** optimizes the previously most efficient method to calculate RPA correlation energies [16, 17] in three ways. First, the accuracy of

the employed integral transforms and, therefore, of the linear-scaling RPA method, is drastically increased by realizing the equivalence of these transforms with the Fourier transform of the non-interacting polarizability, a central quantity in RPA calculations. This optimized transformation not only increases the accuracy, but also reduces the complexity of the calculation by a factor of four, leading to a significant speedup. Second, the memory requirements are reduced by a new scheme to introduce local Cholesky orbitals using solely the ground state one particle density matrix. Finally, a multi-node parallel algorithm is presented that allows to exploit the benefits of modern distributed computing architectures. These developments combined allow to study the layering of covalent organic frameworks, a type of system that was far out of reach of RPA correlation energy calculations beforehand.

As for the study of chemical phenomena access to properties beyond the ground state energy are indispensable, **Publication III** presents a low-scaling method to calculate RPA analytical gradients for large molecular systems. Employing a series of Fourier transforms of the non-interacting polarizability and the correlated screened Coulomb interaction in the calculation of the self-energy to optimally exploit the locality of the respective many-body quantities allows for an asymptotic quadratic scaling analytical gradient method. In comparison to the previously available quartic scaling RI-RPA gradient method [18], this extends the availability of analytical gradients to systems with up to 600 atoms. Due to the optimized quadrature schemes and the use of a compact Cholesky orbital basis, the new method is superior for all system sizes.

While the RPA already significantly increases the accuracy of conventional DFT methods, it has a variety of short comings that can be traced back to the neglect of exchange contributions in the correlation energy. In comparison to the RPA, which is a well defined approximation, it is not straightforward to formulate a ubiquitously applicable RPA with exchange method. The most rigorously derived RPA with exchange methods suffer from the so called triplet instability [19]. Therefore, there is a need for approximate RPA with exchange methods. One popular approach is the second-order screened exchange (SOSEX) method [20–22], which reduces the self-correlation present in RPA correlation energies. In doing so, it however sacrifices the good description of static correlation in stretched molecules of plain RPA [23]. By modifying the SOSEX method to only correct the RPA at short-range distances, a new approximate RPA with exchange scheme is presented in **Publication IV**. The new short-range SOSEX correction results in a more balanced performance of the new method, as exemplified by its superior performance for barrier heights and systems prone to self-interaction error in comparison to both RPA and conventional RPA-SOSEX. Therefore, this method represents a promising new avenue to highly accurate beyond RPA correlation energies.

While the inclusion of exchange effects is crucial to correct the short comings of the RPA, it also drastically increases the computational cost of these methods. While before this thesis RPA with exchange schemes were only accessible to small molecules, linear- and low-scaling methods were devised in **Publication V** that are applicable to almost all different RPA with exchange methods. These methods have extended the realm of RPA with exchange methods to systems containing up

to 500 atoms. This is facilitated by the use of the time-domain representation of the non-interacting response function. Furthermore, Cholesky orbitals and novel integral estimates for the correlated screened Coulomb interaction are introduced. These developments combined allow to test the accuracy of RPA-SOSEX for large, dispersion dominated systems [24], drastically outperforming other electron correlation methods, such as MP2.

The research presented in this thesis significantly contributes to the field of DFT and in particular ACFDT methods. The efficient access to RPA and beyond RPA correlation energies, first order RPA properties and the newly developed short-range RPA-SOSEX method adds to the development of highly accurate alternatives to conventional DFT that are applicable to a wide range of chemical systems.

Chapter 2

Theoretical Basis

A large part of quantum chemistry is concerned with an approximate, computationally efficient solution of the time-independent electronic Schrödinger equation [1] in the Born-Oppenheimer approximation [2]. This chapter first gives a short overview of the different avenues to compute approximate solutions to the electronic Schrödinger equation, which constitute modern electronic structure theory. Then density functional theory and the adiabatic-connection are introduced in more detail. From the adiabatic-connection, the inclusion of exact exchange into DFT is motivated, which defines the field of hybrid DFT. Furthermore, the correlation energy in the adiabatic-connection fluctuation-dissipation theorem is derived, which is the starting point for correlation energies within the random-phase-approximation and beyond.

The remainder of this chapter introduces numerical approximation schemes that allow to lower the computational complexity of expensive correlation energy calculations. In particular it is shown that expressing correlation energies in terms of the non-interacting Green's function, in a compact and local Cholesky-decomposed-density (CDD) atomic orbital (AO) basis in imaginary time allows for low-scaling calculations without computational overhead to the respective canonical theory.

The combination of these topics forms the basis for the research on DFT and ACFDT methods presented in the next chapter.

2.1 Basis of Modern Electronic Structure Theory

The numerically exact solution of the electronic, time-independent Schrödinger equation in the Born-Oppenheimer approximation [2]

$$\begin{aligned} \hat{H}\left|\Psi_n\right\rangle &= E_n\left|\Psi_n\right\rangle, & (2.1)\\ \hat{H} &= \hat{T}_{\mathrm{el}} + \hat{V}_{\mathrm{eN}} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{NN}}, & (2.2) \end{aligned}$$

with the Ritz method [25] is commonly referred to as FCI (for a review see, e.g., Ref. [26]). In this approach the wave function is expanded in the complete set of Slater-Determinants

$$\left|\Psi_0^{\mathrm{FCI}}\right\rangle = \sum_i c_i \left|\Phi_i\right\rangle \tag{2.3}$$

within a given basis set. With this ansatz, Eq. 2.1 is solved by diagonalization. Since the computational cost of this approach scales exponentially with the basis set size, it is limited to light atoms and the smallest molecules. Therefore, approximate yet accurate methods are required to tackle the many-body problem. This research question defines the field of electronic structure theory. In the following, the milestones that lay the foundation of modern quantum chemistry are briefly outlined.

2.1.1 Hartree-Fock Theory

The Hartree-Fock method [3–5] was historically one of the first successful approximation schemes, and is still widely used today, also as a starting point for more accurate electronic structure methods. It is defined by the use of a single Slater-Determinant (SD) of one-electron molecular orbitals (MOs) with N_{el} (number of electrons) spatial-spin coordinates $x_i = (r_i, \sigma_i)$ as an ansatz for the wave function

$$\Phi_0^{\mathrm{SD}}(x_1, x_2, \ldots, x_{N_{\mathrm{el}}}) = \frac{1}{\sqrt{N_{\mathrm{el}}}} \begin{vmatrix} \varphi_1(x_1) & \varphi_1(x_2) & \ldots & \varphi_1(x_{N_{\mathrm{el}}}) \\ \varphi_2(x_1) & \ldots & & \ldots \\ \ldots & & \ldots & \ldots \\ \varphi_{N_{\mathrm{el}}}(x_1) & \ldots & \ldots & \varphi_{N_{\mathrm{el}}}(x_{N_{\mathrm{el}}}) \end{vmatrix}. \tag{2.4}$$

Variational minimization of the expectation value of the Hamiltonian

$$E_0^{\mathrm{HF}} = \frac{\langle \Phi_0^{\mathrm{SD}} | \, \hat{H} \, | \Phi_0^{\mathrm{SD}} \rangle}{\langle \Phi_0^{\mathrm{SD}} | \Phi_0^{\mathrm{SD}} \rangle} \tag{2.5}$$

with respect to the one-electron functions in the SD under the orthonormality constraint of the MOs leads to the well known canonical HF equations

$$\hat{F}\varphi_i(x_1) = \epsilon_i \varphi_i(x_1), \tag{2.6}$$

$$\hat{F} = \hat{T} + \hat{V}_{\mathrm{N}} + \hat{J} + \hat{K}, \tag{2.7}$$

$$\hat{T} = -\frac{\vec{\nabla}^2}{2}, \tag{2.8}$$

$$\hat{V}_{\mathrm{N}} = -\sum_{A}^{N_{\mathrm{N}}} \frac{Z_A}{r_{1A}}, \tag{2.9}$$

$$\hat{J} = \sum_{j}^{N_{\mathrm{el}}} \int \mathrm{d}x_2 \varphi_j^*(x_2) \frac{1}{r_{12}} \varphi_j(x_2), \tag{2.10}$$

$$\hat{K} = -\sum_{j}^{N_{\mathrm{el}}} \int \mathrm{d}x_2 \varphi_j^*(x_2) \frac{\hat{P}}{r_{12}} \varphi_j(x_2) \tag{2.11}$$

for a system with N_{N} nuclei with charges Z_A. Atomic units are used throughout this thesis. Eq. 2.6 can be solved routinely with the introduction of basis functions in the Roothan-Hall scheme [27, 28]. The most prominent families of basis functions in modern electronic structure theory are Gaussian type orbitals [29], which are found mainly in the field of molecular quantum chemistry, and plane-waves in solid-state physics. The HF ansatz is the simplest method which captures the correct anti-symmetry of an electronic wave function. With this, it also includes

non-classical same-spin electronic correlation effects, which are referred to as exchange interactions. The exchange-interactions manifest in the exchange-operator $\hat{K}$, where $\hat{P}$ permutes the coordinate in the operator with the coordinate in the function it acts on. As the HF equations are, however, single particle equations, the electron-electron interaction is only described at the mean-field level through the Coulomb $\hat{J}$ and exchange $\hat{K}$ operators.

2.1.2 Electron Correlation

While the HF approach captures the majority of the FCI total energies, it typically fails to achieve quantitatively accurate results for observable quantities. The remaining part of the energy that is missing in the HF description was defined by Löwdin [30] as electron correlation. It is one of the biggest open research questions in quantum chemistry to devise computationally efficient approaches to describe electron correlation.
The conventional avenue to tackle the electron correlation problem are so called wave function based methods, among which the most popular textbook examples are Møller-Plesset second-order perturbation theory (MP2) [31] and Coupled-Cluster [32, 33] theory. The Coupled-Cluster (CC) ansatz, which is defined by an exponential parameterization of the wave function in terms of excitations with reference to the HF-SD, leads systematically to the FCI result if all excitations levels are incorporated. As this again leads to exponential computational requirements, practical CC calculations truncate the excitation level, which leads to polynomial scaling algorithms such as CC with single and double excitations (CCSD). Here, one trades computational efficiency for accuracy. Over the years, the CCSD method including perturbative triples (CCSD(T)) [34], established itself as the so called gold standard of quantum chemistry, since it delivers chemically accurate (≈ 1 kcal/mol) results for a wide range of problems. In its canonical formulation CCSD(T) scales as $\mathcal{O}(N^7)$ with the molecule size N, which limits its applicability to small and medium sized systems.
The electron correlation present in the type of chemical systems for which CCSD(T) and CC or MP2 methods in general deliver accurate results is dominated by dynamic correlation effects. Dynamic correlation arises from electron scattering effects, which are missing in the mean-field treatment of the electron-electron interaction in HF. Static correlation on the other hand describes correlation effects which originate from the fact that in these cases the wave function requires a multi-determinantal description. For systems dominated by static correlation, all methods that start from one determinant are usually inadequate, and one has to ressort to multiconfigurational methods [35] such as complete-active-space self-consistent-field (CAS-SCF) calculations, potentially with an on top dynamic correlation treatment such as CAS second-order perturbation theory (CAS-PT2) [36, 37].
In their canonical formulation all the methods described in this section scale at least as $\mathcal{O}(N^5)$ with respect to the system size. Furthermore, correlated calculations show slow convergence with the basis set size (see, e.g., Ref. [38]). These two facts combined explain the steep computational scaling of wave function based correlation methods, which hamper their widespread application. A variety of techniques exist to accelarate the canonical algorithms of abovementioned theories, some of which will be discussed in Section 2.5. Nevertheless, tackling large molecular systems with

significantly more than 1000 atoms, or calculating multiple thousands of energies, as, e.g., required in molecular dynamics simulations, with wave function based correlation methods is not yet possible with the computing resources typically available to quantum chemistry researchers. For these cases, the typical method of choice is DFT, which has similar computational complexity as HF, includes, however, an approximate description of electron correlation.

2.1.3 Density Functional Theory

An alternative approach to wave function theories is density functional theory. The formulation of DFT nowadays is based on the work of Hohenberg, Kohn and Sham. The foundations of DFT were laid by the Hohenberg-Kohn theorems in 1964 [39]. The first theorem states that there is a one to one correspondence between the ground state density and the external potential $V_{\mathrm{ext}}(r)$. Since the ground state density determines the number of electrons, it also fixes the ground state wave function and all ground state properties. Thus, the ground state energy is a functional of the ground state density, for which the second Hohenberg-Kohn theorem states a variational principle. In principle one could, therefore, solve for the ground state energy through functional minimization of

$$E[n] = \int \mathrm{d}r n(r) V_{\mathrm{ext}}(r) + F^{\mathrm{HK}}[n], \tag{2.12}$$

$$n(r_1) = \int \ldots \int \mathrm{d}\sigma_1 \mathrm{d}x_2 \ldots \mathrm{d}x_{N_{\mathrm{el}}} |\Psi_0(x_1, x_2, \ldots, x_{N_{\mathrm{el}}})|^2, \tag{2.13}$$

where $F^{\mathrm{HK}}[n]$ is the unknown Hohenberg-Kohn functional and $n(r)$ the ground-state density. For the discussion of $N-, v-$representability and the constrained search formalism the reader is referred to a standard textbook on the subject [40]. As the ground state density is a much simpler object than the N-electron wave function, this might lead to a computationally significantly less involved procedure to determine the ground state energy. The caveat is, however, that the exact Hohenberg-Kohn functional is unknown and approximations to the exact functional are necessary. Basically all of the nowadays highly succesful DFT approaches are based on the Kohn-Sham (KS) formulation of DFT [41]. The idea of Kohn and Sham was to introduce a non-interacting SD $|\Phi_0^{\mathrm{KS}}\rangle$, which yields the same ground state density as the interacting system and is determined through minimization of the functional

$$E[n] = T_{\mathrm{s}}[n] + E_{\mathrm{H}}[n] + E_{\mathrm{ext}}[n] + E_{\mathrm{xc}}[n], \tag{2.14}$$

$$T_{\mathrm{s}}[n] = -\sum_i^{N_{\mathrm{el}}} \int \mathrm{d}x \varphi_i^*(x) \frac{\vec{\nabla}^2}{2} \varphi_i(x), \tag{2.15}$$

$$E_{\mathrm{H}}[n] = \frac{1}{2} \int \int \mathrm{d}r_1 \mathrm{d}r_2 \frac{n(r_1) n(r_2)}{r_{12}}, \tag{2.16}$$

$$E_{\mathrm{ext}}[n] = \int \mathrm{d}r n(r) V_{\mathrm{ext}}(r), \tag{2.17}$$

$$E_{\mathrm{xc}}[n] = T[n] - T_{\mathrm{s}}[n] + V_{\mathrm{ee}}[n] - E_{\mathrm{H}}[n], \tag{2.18}$$

with an orthonormality constraint for the MOs. This has the advantage that most of the kinetic energy (Eq. 2.15) of the system and the Hartree repulsion (Eq. 2.16)

can be represented exactly and only the missing part of the kinetic energy ($T[n] - T_\mathrm{s}[n]$) and the electron-electron interaction ($V_\mathrm{ee}[n] - E_\mathrm{H}[n]$) have to be approximated through $E_\mathrm{xc}[n]$. minimization of Eq. 2.14 with an orthonomality constraint for the MOs leads to the following system of one-particle equations analogous to the canonical HF equations (Eq. 2.6)

$$\hat{h}_\mathrm{KS}\varphi_i(x_1) = \epsilon_i\varphi_i(x_1). \tag{2.19}$$

In case the external potential V_ext matches the nuclear potential V_N, the one-particle Kohn-Sham Hamiltonian is given as

$$\begin{aligned}
\hat{h}_\mathrm{KS} &= \hat{T} + \hat{V}_\mathrm{KS}, && (2.20)\\
\hat{V}_\mathrm{KS} &= \hat{V}_\mathrm{N} + \hat{V}_J + \hat{V}_\mathrm{xc} = V_\mathrm{KS}(r), && (2.21)\\
\hat{V}_\mathrm{J} &= \int \mathrm{d}r_2 \frac{n(r_2)}{r_{12}}, && (2.22)\\
\hat{V}_\mathrm{xc} &= \frac{\partial E_\mathrm{xc}[n(r)]}{\partial n(r)} = V_\mathrm{xc}(r), && (2.23)
\end{aligned}$$

where the local and multiplicative exchange-correlation potential V_xc is given as the functional derivative of the exchange-correlation functional, which still needs to be approximated. In this formulation, DFT promises an exact solution to the ground-state Schrödinger equation with similar computational complexity as the HF scheme, if the exchange-correlation potential were known.

2.1.4 Jacob's-Ladder of Density Functional Theory

The Jacob's-Ladder of DFT introduced by Perdew [42] describes a hierarchy of DFAs, where DFAs on a higher rung incorporate more ingredients into the approximate exchange-correlation functional. Going up Jacob's ladder therefore leads to increased computational cost, but ideally also to more accurate results. In total, there are five rungs on Jacob's ladder, the lowest one being functionals that solely take the electron density $n(r)$ as an input. These functionals are known as local density approximations (LDA). The next rung of DFAs is known as generalized-gradient-approximations (GGA), which in addition to the LDA include the gradient of the density. The third rung functionals include occupied orbitals in the form of, e.g., the kinetic-energy gradient and the fourth rung in form of a fraction of exact exchange (Eq. 2.34). DFAs of the third rung are known as meta-GGA functionals and the fourth rung as hybrid functionals. As meta-GGAs are, strictly speaking, implicit density functionals due to the appearance of MOs and hybrid functionals do not employ local, multiplicative exchange-correlation potentials, they fall in the realm of the generalized Kohn-Sham (GKS) scheme [43]. Finally the fifth rung is defined by the use of virtual orbitals. This can be in the form of, e.g., a second-order perturbation correction (PT2) [44] or random-phase-approximation (RPA) correlation energies, which are typically calculated as a post KS correction to a preceding self-consistent DFT calculation with one of the DFAs belonging to rungs I-IV.
The research presented in this thesis extends and optimizes the computational efficiency of DFAs belonging to the fourth and fifth rung of Jacob's ladder. As the adiabatic-connection can be seen as the basis for both hybrid DFT and ACFDT correlation energies it is derived in the next section. Section 2.3 will go into more

detail of hybrid DFT and especially range-separated hybrid DFT calculations, which is the topic of **Publication I**. Then the RPA and beyond RPA approaches within the ACFDT framework are introduced in Section 2.4, which are the subject of **Publications II-V**.

2.2 Adiabatic-Connection

The adiabatic-connection is a very important and successful concept in DFT for the derivation of exchange-correlation functionals. While the AC was first introduced by Gunnarson and Lundqvist [9] and Langreth and Perdew [10, 11], the derivation here follows the textbook approach of Engel and Dreizler [45]. The AC relies on a coupling-strength Hamiltonian

$$\hat{H}^\lambda = \hat{T} + \lambda \hat{V}_{\mathrm{ee}} + \hat{V}^\lambda, \tag{2.24}$$

where $\hat{V}_{\mathrm{ee}}$ is the exact electron-electron interaction and $\hat{V}^\lambda$ consists of a local, multiplicative potential. The coupling-strength parameter λ transforms the Hamiltonian from the non-interacting case at $\lambda = 0$ to the fully-interacting case at $\lambda = 1$. The local, multiplicative potential $\hat{V}^\lambda$ is defined such that the coupling strength Hamiltonian $\hat{H}^\lambda$ yields the same ground state density n for each value of λ and therefore reduces to the external potential $\hat{V}_{\mathrm{ext}}$ ($\hat{V}_{\mathrm{N}}$, Eq. 2.9, in the absence of another field) for $\lambda = 1$ and to $\hat{V}_{\mathrm{KS}}$ for $\lambda = 0$. Furthermore, the normalized, coupling-strength dependent ground state Ψ_0^λ is given as the eigenfunction of $\hat{H}^\lambda$

$$\hat{H}^\lambda \left|\Psi_0^\lambda\right\rangle = E_0^\lambda \left|\Psi_0^\lambda\right\rangle, \tag{2.25}$$

$$1 = \left\langle \Psi_0^\lambda | \Psi_0^\lambda \right\rangle. \tag{2.26}$$

The interacting ground state is given as $\Psi_0^{\lambda=1}$ and $\Psi_0^{\lambda=0} = \Phi_0^{\mathrm{KS}}$ corresponds to the non-interacting KS-SD. Now, a formula for the interacting ground state energy can be derived by employing the identity

$$E_0^1 - E_0^0 = \int_0^1 \mathrm{d}\lambda \frac{\partial}{\partial \lambda} \left\langle \Psi_0^\lambda \right| \hat{H}^\lambda \left|\Psi_0^\lambda\right\rangle. \tag{2.27}$$

Using the Hellman-Feynman theorem [46], which is valid along the coupling-strength path gives

$$E_0^1 - E_0^0 = \int_0^1 \mathrm{d}\lambda \left\langle \Psi_0^\lambda \right| \hat{V}_{\mathrm{ee}} + \frac{\partial \hat{V}^\lambda}{\partial \lambda} \left|\Psi_0^\lambda\right\rangle \tag{2.28}$$

$$= \int_0^1 \mathrm{d}\lambda \left\langle \Psi_0^\lambda \right| \hat{V}_{\mathrm{ee}} \left|\Psi_0^\lambda\right\rangle + \int_0^1 \mathrm{d}\lambda \int \mathrm{d}r n(r) \frac{\partial V^\lambda(r)}{\partial \lambda} \tag{2.29}$$

Inserting the limits for $\hat{V}^\lambda$ at $\lambda = 0$ and $\lambda = 1$ allows to write

$$E_0^1 - E_0^0 = \int_0^1 \mathrm{d}\lambda \left\langle \Psi_0^\lambda \right| \hat{V}_{\mathrm{ee}} \left|\Psi_0^\lambda\right\rangle + \int \mathrm{d}r n(r)[V_{\mathrm{ext}}(r) - V_{\mathrm{KS}}(r)]. \tag{2.30}$$

With the non-interacting energy expression

$$E_0^0 = \left\langle \Phi_0^{\mathrm{KS}} \right| \hat{H}^0 \left|\Phi_0^{\mathrm{KS}}\right\rangle = \left\langle \Phi_0^{\mathrm{KS}} \right| \hat{T} \left|\Phi_0^{\mathrm{KS}}\right\rangle + \int \mathrm{d}r n(r) V_{\mathrm{KS}}(r), \tag{2.31}$$

the ground state energy can be written as

$$E_0^1 = \langle \Phi_0^{\mathrm{KS}} | \hat{T} | \Phi_0^{\mathrm{KS}} \rangle + \int \mathrm{d}r n(r) V_{\mathrm{ext}}(r) + \int_0^1 \mathrm{d}\lambda \langle \Psi_0^\lambda | \hat{V}_{\mathrm{ee}} | \Psi_0^\lambda \rangle . \tag{2.32}$$

Comparing Eq. 2.32 with Eq. 2.14 allows to identify the exchange-correlation energy E_{xc} after subtracting the Hartree-repulsion (Eq. 2.16)

$$E_{\mathrm{xc}} = \int_0^1 \mathrm{d}\lambda \langle \Psi_0^\lambda | \hat{V}_{\mathrm{ee}} | \Psi_0^\lambda \rangle - \frac{1}{2} \int\int \mathrm{d}r_1 \mathrm{d}r_2 \frac{n(r_1) n(r_2)}{r_{12}}. \tag{2.33}$$

The result in Eq. 2.33 serves as the starting point for a variety of DFAs [47–49]. For the construction of LDA and GGA functionals one usually evaluates Eq. 2.33 for the model systems of the homogenuous electron gas or the slowly varying electron gas to construct exchange-correlation functionals, which depend only on the density or the gradient of the density. One popular example for such a functional is, e.g., the VWN correlation functional [47].
Eq. 2.33 is also the foundation of all the research presented in this thesis. If one expands the coupling-strength integrand in λ in the non-interacting limit (i.e., $\lambda \rightarrow 0$) one obtains the exact exchange energy E_{X} expression in terms of KS orbitals

$$E_{\mathrm{X}} = -\frac{1}{2} \sum_{i,j}^{N_{\mathrm{el}}} \int\int \mathrm{d}x_1 \mathrm{d}x_2 \frac{\varphi_i^*(r_1) \varphi_j^*(r_2) \varphi_i(r_2) \varphi_j(r_1)}{r_{12}}, \tag{2.34}$$

as the first non-vanishing term. This was the motivation of Becke [48] to establish the class of hybrid DFT functionals including a fraction of KS exact exchange. Range-separated hybrid DFT is a special case of hybrid DFT where the admixture of exchange is different at different length scales. **Publication I** is concerned with short-range hybrid functionals, where short-range exact exchange is employed together with long range KS exchange. More details are given in Section 2.3
Furthermore, Eq. 2.33 is the basis of all correlation energy formulas used in the context of RPA and RPA with exchange schemes, where the coupling-strength integrand is evaluated with the use of approximate response functions of the system at hand. This is the topic of Section 2.4.

2.3 Hybrid Density Functional Theory

While pure KS-DFT has produced very successful DFAs such as, e.g., the Perdew-Burke-Ernzerhof (PBE) functional [50], it has its limitations. The problems of pure DFT can, e.g., be seen in the relatively weak performance for thermochemistry [48]. These problems have motivated to include a fraction of exact exchange in DFT calculations. A physical motivation from first principles can be given by use of the AC [48], which was introduced in the previous section. This initiated the field of hybrid DFT. The general formula for the total energy in hybrid DFT reads

$$E = T_{\mathrm{s}} + E_{\mathrm{ext}} + E_{\mathrm{H}} + a E_{\mathrm{X}}^{\mathrm{KS}} + b E_{\mathrm{X}} + E_{\mathrm{C}}^{\mathrm{KS}}. \tag{2.35}$$

The factors a and b vary between functionals. Perdew et al. [51] used arguments from perturbation theory for setting $b = 0.25$, which is the basis for the famous

PBE0 functional [52]. While Eq. 2.35 allows to combine DFT exchange and exact exchange it might be advantagous to have a more flexible combination of the two at different length scales. This is facilitated by range-separated DFT [53]. One avenue to range-separated DFT is the use of Coulomb attenuation. Here, one separates the Coulomb operator in a short-range (sr) and a long-range (lr) component. One possible way to do this is through use of the error function [54]

$$\frac{1}{r_{12}} = \frac{\mathrm{erf}(wr_{12})}{r_{12}} + \frac{\mathrm{erfc}(wr_{12})}{r_{12}}, \tag{2.36}$$

where the error function approaches unity for large wr_{12} and goes to zero for small arguments. The complementary error function adds the remaining part to unity and therefore shows exactly the opposite behavior. Therefore, the first term in Eq. 2.36 describes long-range Coulomb interactions while the second term describes short-range Coulomb effects. This is then used to seperate exact exchange E_X and KS exchange E_X^KS into a short-range and long-range component. For the exact exchange energy the expressions are readily obtained as

$$E_\mathrm{X} = E_\mathrm{X}^{sr}(w) + E_\mathrm{X}^{lr}(w), \tag{2.37}$$

$$E_\mathrm{X}^{sr}(w) = -\frac{1}{2}\sum_{i,j}^{N_\mathrm{el}} \int\int \mathrm{d}x_1 \mathrm{d}x_2 \frac{\varphi_i^*(r_1)\varphi_j^*(r_2)\mathrm{erfc}(wr_{12})\varphi_i(r_2)\varphi_j(r_1)}{r_{12}}, \tag{2.38}$$

$$E_\mathrm{X}^{lr}(w) = -\frac{1}{2}\sum_{i,j}^{N_\mathrm{el}} \int\int \mathrm{d}x_1 \mathrm{d}x_2 \frac{\varphi_i^*(r_1)\varphi_j^*(r_2)\mathrm{erf}(wr_{12})\varphi_i(r_2)\varphi_j(r_1)}{r_{12}}. \tag{2.39}$$

For the KS exchange part this separation has to be performed individually for the respective exchange DFT functional as, e.g., for short-range PBE in Ref. [55]. This gives rise to a more flexible formula for the range-separated hybrid DFT total energy

$$E = T_\mathrm{s} + E_\mathrm{ext} + E_\mathrm{H} + aE_\mathrm{X}^{\mathrm{KS},sr} + bE_\mathrm{X}^{\mathrm{KS},lr} + cE_\mathrm{X}^{sr} + dE_\mathrm{X}^{lr} + E_\mathrm{C}^\mathrm{KS}. \tag{2.40}$$

Range-separated DFT calculations which mainly use exact exchange in the long-range regime are typically used to model charge-transfer phenomena such as it is done in the CAM-B3LYP DFA [56]. Short-range hybrid DFT sets $d = 0$ and only keeps the short-range exact exchange fraction while modeling the rest with KS exchange. Here, the most-popular example is the HSE06 functional [57–60]. The motivation for short-range hybrid DFT is that while including exact exchange into DFT calculations typically increases the accuracy, it also drastically increases the computational cost. Especially in periodic systems this can lead to significant problems due to convergence problems of exact exchange with the number of unit cells, even though there are new developments trying to circumvent this problem [61]. But also in molecular calculations exact exchange is the time determining step in SCF calculations for large systems. Here, short-range DFT can be a compromise between including portions of exact exchange for increased accuracy while keeping the computational cost low. **Publication I** presents integral screening methods to exploit the short-range nature of short-range exchange in linear-scaling methods to calculate the exchange matrix. This makes short-range hybrid DFT calculation computationally similar in cost to pure hybrid DFT calculations while retaining the gain in accuracy.

2.4 Adiabatic-Connection Fluctuation-Dissipation Theorem

Based on the AC one can not only motivate the use of exact exchange, but also derive an exact formula for the correlation energy E_C in terms of density response functions with the help of the adiabatic-connection fluctuation-dissipation theorem. The random-phase-approximation [12] is then the simplest approximation that turns the ACFDT expression into a practical theory for molecular calculations. As the RPA, however, still suffers from a variety of problems one can further include exchange effects into RPA correlation energies to improve upon the RPA. In the following, the working equations that build the basis for **Publications II-V** are derived. As in the AC section the derivation follows the text book approach of Engel and Dreizler [45].

2.4.1 Correlation Energy in Terms of Response Functions

To obtain an expression for the correlation energy E_C from Eq. 2.33 one starts with its definition in the context of AC DFT

$$E_\mathrm{C} = E_\mathrm{xc} - E_\mathrm{X}, \tag{2.41}$$

where E_X is given by Eq. 2.34. Realizing that the expectation value of $\hat{V}_\mathrm{ee}$ with a Slater-determinant is given as the sum of the Hartree- (Eq. 2.16) and the exact exchange energy (Eq. 2.34) as well known from HF theory

$$\langle \Phi_0^\mathrm{KS} | \, \hat{V}_\mathrm{ee} \, | \Phi_0^\mathrm{KS} \rangle = \frac{1}{2} \int \int \mathrm{d}r_1 \mathrm{d}r_2 \frac{n(r_1) n(r_2)}{r_{12}} + E_\mathrm{X} \tag{2.42}$$

one obtains for Eq. 2.41

$$E_\mathrm{C} = \int_0^1 \mathrm{d}\lambda \, \langle \Psi_0^\lambda | \, \hat{V}_\mathrm{ee} \, | \Psi_0^\lambda \rangle - \langle \Phi_0^\mathrm{KS} | \, \hat{V}_\mathrm{ee} \, | \Phi_0^\mathrm{KS} \rangle \tag{2.43}$$

as an exact formula for the correlation energy.

Introduction of Density-Fluctuation Operators

To evaluate Eq. 2.43 it is convenient to work in second quantization. In second quantization one employs the fermionic creation $\hat{a}_p^\dagger$ and annihilation operators $\hat{a}_p$, creating or destroying a particle in the one-particle state p. To incorporate the correct fermion statistics they obey the following anti-commutation relations

$$\{\hat{a}_p^\dagger, \hat{a}_q^\dagger\} = 0, \tag{2.44}$$
$$\{\hat{a}_p, \hat{a}_q\} = 0, \tag{2.45}$$
$$\{\hat{a}_p^\dagger, \hat{a}_q\} = \delta_{pq}. \tag{2.46}$$

Here, δ_{pq} represents the Kronecker delta. Alternatively, one can employ the real-space field operators $\hat{\Psi}^\dagger(x), \hat{\Psi}(x)$ which respectively create or delete a particle at

$x = (r, \sigma)$. One can switch from one representation to the other using that the one-particle basis forms a complete set in the one-particle Hilbert space

$$\hat{\Psi}(x) = \sum_p \varphi_p(x)\hat{a}_p, \tag{2.47}$$

$$\hat{a}_p = \int \mathrm{d}x \varphi_p(x)\hat{\Psi}(x), \tag{2.48}$$

and the analogous formulas for the creation operators. From Eqs. 2.47 and 2.48 result the anti-commutation rules for the field operators

$$\{\hat{\Psi}^\dagger(x_1), \hat{\Psi}^\dagger(x_2)\} = 0, \tag{2.49}$$

$$\{\hat{\Psi}(x_1), \hat{\Psi}(x_2)\} = 0, \tag{2.50}$$

$$\{\hat{\Psi}^\dagger(x_1), \hat{\Psi}(x_2)\} = \delta_{x_1,x_2}. \tag{2.51}$$

In second quantization the electron-electron interaction is now given as

$$\hat{V}_{\mathrm{ee}} = \frac{1}{2}\int\int \mathrm{d}x_1 \mathrm{d}x_2 \frac{\hat{\Psi}^\dagger(x_1)\hat{\Psi}^\dagger(x_2)\hat{\Psi}(x_2)\hat{\Psi}(x_1)}{r_{12}} \tag{2.52}$$

where the field operators can be rearranged using Eqs. 2.49, 2.50 and 2.51

$$\hat{\Psi}^\dagger(x_1)\hat{\Psi}^\dagger(x_2)\hat{\Psi}(x_2)\hat{\Psi}(x_1) = \hat{\Psi}^\dagger(x_1)\hat{\Psi}(x_1)\hat{\Psi}^\dagger(x_2)\hat{\Psi}(x_2) - \delta_{x_1,x_2}\hat{\Psi}^\dagger(x_1)\hat{\Psi}(x_1). \tag{2.53}$$

Now the density operator is introduced

$$\hat{n}(r) = \int \mathrm{d}\sigma \hat{\Psi}^\dagger(r\sigma)\hat{\Psi}(r\sigma) \tag{2.54}$$

to rewrite $\hat{V}_{\mathrm{ee}}$ as

$$\hat{V}_{\mathrm{ee}} = \frac{1}{2}\int\int \mathrm{d}r_1 \mathrm{d}r_2 v(r_1, r_2)[\hat{n}(r_1)\hat{n}(r_2) - \delta_{r_1 r_2}\hat{n}(r_1)], \tag{2.55}$$

where $v(r_1, r_2) = \frac{1}{r_{12}}$ for convenience of notation. Therefore, Eq. 2.43 can be transformed to

$$\begin{aligned} E_{\mathrm{C}} = \frac{1}{2}\int_0^1 \mathrm{d}\lambda \int\int \mathrm{d}r_1 \mathrm{d}r_2 v(r_1, r_2)[&\langle\Psi_0^\lambda|\, \hat{n}(r_1)\hat{n}(r_2) - \delta_{r_1 r_2}\hat{n}(r_1)\, |\Psi_0^\lambda\rangle \\ &- \langle\Phi_0^{\mathrm{KS}}|\, \hat{n}(r_1)\hat{n}(r_2) - \delta_{r_1 r_2}\hat{n}(r_1)\, |\Phi_0^{\mathrm{KS}}\rangle]. \end{aligned} \tag{2.56}$$

As per construction, the expectation value of the density operator yields the ground state density along the adiabatic-connection path

$$n(r) = \langle\Psi_0^\lambda|\, \hat{n}(r)\, |\Psi_0^\lambda\rangle\,, \tag{2.57}$$

the density-fluctuation operators can be introduced

$$\hat{\tilde{n}}(r) = \hat{n}(r) - n(r) \tag{2.58}$$

to simplify Eq. 2.56

$$\begin{aligned} E_{\mathrm{C}} = \frac{1}{2}\int_0^1 \mathrm{d}\lambda \int\int \mathrm{d}r_1 \mathrm{d}r_2 v(r_1, r_2)[&\langle\Psi_0^\lambda|\, \hat{\tilde{n}}(r_1)\hat{\tilde{n}}(r_2)\, |\Psi_0^\lambda\rangle \\ &- \langle\Phi_0^{\mathrm{KS}}|\, \hat{\tilde{n}}(r_1)\hat{\tilde{n}}(r_2)\, |\Phi_0^{\mathrm{KS}}\rangle]. \end{aligned} \tag{2.59}$$

Introduction of the Response Function

To transform Eq. 2.59 into working equations one uses the resemblance of the expectation values of the product of the density-fluctuation operators with the time-ordered response function, a well known quantity in many-body perturbation theory (MBPT) and time-dependent density functional theory (TD-DFT):

$$\chi(r_1t_1, r_2t_2) = -\mathrm{i} \left\langle \Psi_0 \right| T[\hat{\tilde{n}}(r_1t_1)\hat{\tilde{n}}(r_2t_2)] \left| \Psi_0 \right\rangle . \tag{2.60}$$

Here, the density-fluctuation operator in the Heisenberg picture is given as

$$\hat{\tilde{n}}(rt) = \hat{n}(rt) - n(r), \tag{2.61}$$

$$\hat{n}(rt) = e^{\mathrm{i}\hat{H}t}\hat{n}(r)e^{-\mathrm{i}\hat{H}t}, \tag{2.62}$$

and T represents the time-ordering symbol with its usual definition [62]. The response function along the coupling-strength path will be referred to as χ_λ and the coupling-strength index is omitted for general discussion of the response function. Comparing Eq. 2.59 with Eq. 2.60 shows that in this case only the limit $t_1, t_2 \to 0$ is of interest

$$E_\mathrm{C} = \frac{1}{2}\int_0^1 \mathrm{d}\lambda \int\int \mathrm{d}r_1\mathrm{d}r_2 v(r_1, r_2)\{\mathrm{i}\lim_{t_1,t_2\to 0}[\chi_\lambda(r_1t_1, r_2t_2) - \chi_0(r_1t_1, r_2t_2)]\}. \tag{2.63}$$

Lehmann Representation of the Response Function

Inserting the definition of the density-fluctuation operator (Eq. 2.61) in Eq. 2.60 yields

$$\chi(r_1t_1, r_2t_2) = -\mathrm{i}[\left\langle \Psi_0 \right| T[\hat{n}(r_1t_1)\hat{n}(r_2t_2)] \left| \Psi_0 \right\rangle - n(r_1)n(r_2)]. \tag{2.64}$$

To obtain a convenient expression for the response function, the time ordering is given explicitly and a complete set of states $\sum_n \left| \Psi_n \right\rangle \left\langle \Psi_n \right| = \hat{1}$ is inserted between the density operators in Eq. 2.64

$$\begin{aligned}\chi(r_1t_1, r_2t_2) = -\mathrm{i}\Big\{ \sum_n [&\theta(t_1 - t_2) \left\langle \Psi_0 \right| \hat{n}(r_1t_1) \left| \Psi_n \right\rangle \left\langle \Psi_n \right| \hat{n}(r_2t_2) \left| \Psi_0 \right\rangle \\ &- \theta(t_2 - t_1) \left\langle \Psi_0 \right| \hat{n}(r_2t_2) \left| \Psi_n \right\rangle \left\langle \Psi_n \right| \hat{n}(r_1t_1) \left| \Psi_0 \right\rangle] \\ &- n(r_1)n(r_2)\Big\},\end{aligned} \tag{2.65}$$

with the Heaviside step function $\theta(t_1 - t_2)$. This factorizes the expectation value of the product of density operators:

$$\begin{aligned}&\left\langle \Psi_0 \right| \hat{n}(r_1t_1) \left| \Psi_n \right\rangle \left\langle \Psi_n \right| \hat{n}(r_2t_2) \left| \Psi_0 \right\rangle \\ &\qquad = \left\langle \Psi_0 \right| e^{\mathrm{i}\hat{H}t_1}\hat{n}(r_1)e^{-\mathrm{i}\hat{H}t_1} \left| \Psi_n \right\rangle \left\langle \Psi_n \right| e^{\mathrm{i}\hat{H}t_2}\hat{n}(r_2)e^{-\mathrm{i}\hat{H}t_2} \left| \Psi_0 \right\rangle \\ &\qquad = \left\langle \Psi_0 \right| \hat{n}(r_1) \left| \Psi_n \right\rangle \left\langle \Psi_n \right| \hat{n}(r_2) \left| \Psi_0 \right\rangle e^{\mathrm{i}E(t_1-t_2)}e^{-\mathrm{i}E_n(t_1-t_2)}.\end{aligned} \tag{2.66}$$

Eq. 2.66 shows that the time-ordered response function depends only on the time difference $(t_1 - t_2)$ and can therefore be expressed more conveniently in frequency space. Fourier transformation leads to the so called Lehmann representation [63] of

the response function

$$\chi(r_1, r_2, \omega) = \int_{-\infty}^{\infty} \mathrm{d}(t_1 - t_2) e^{\mathrm{i}\omega(t_1 - t_2)} \chi(r_1, r_2, t_1 - t_2), \tag{2.67}$$

$$\begin{aligned} \chi(r_1, r_2, \omega) = & \sum_{n \neq 0} \frac{\langle \Psi_0 | \hat{n}(r_1) | \Psi_n \rangle \langle \Psi_n | \hat{n}(r_2) | \Psi_0 \rangle}{\omega - (E_n - E_0) + \mathrm{i}\eta} \\ & - \sum_{n \neq 0} \frac{\langle \Psi_0 | \hat{n}(r_2) | \Psi_n \rangle \langle \Psi_n | \hat{n}(r_1) | \Psi_0 \rangle}{\omega + (E_n - E_0) - \mathrm{i}\eta}, \end{aligned} \tag{2.68}$$

where the ground state $\Psi_{n=0}$ in the sum over states cancels the product of ground state densities in Eq. 2.65. The infinitesimal pole shifts $\mathrm{i}\eta$ with $\eta > 0$ arise from the integral formulation of the Heaviside step function. Now the frequency-dependent response function is used to express the $t_1, t_2 \to 0$ limit in Eq. 2.63

$$\begin{aligned} \lim_{t_1, t_2 \to 0} \chi(r_1, r_2, t_1 - t_2) &= \lim_{t_1, t_2 \to 0} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} e^{-\mathrm{i}\omega(t_1 - t_2)} \chi(r_1, r_2, \omega) \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \chi(r_1, r_2, \omega). \end{aligned} \tag{2.69}$$

Since Eqs. 2.65, 2.66, 2.67 and 2.68 hold for any Ψ_0^λ and $\hat{H}^\lambda$ along the AC, Eq. 2.63 can be expressed as

$$E_\mathrm{C} = \frac{1}{2} \int_0^1 \mathrm{d}\lambda \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \int\int \mathrm{d}r_1 \mathrm{d}r_2 v(r_1, r_2) \mathrm{i}[\chi_\lambda(r_1, r_2, \omega) - \chi_0(r_1, r_2, \omega)]. \tag{2.70}$$

The Lehmann representation of the response function (Eq. 2.68) shows that $\chi(r_1, r_2, \omega)$ is symmetric with respect to the frequency ω. Therefore, the ω-integration in Eq. 2.70 can be restricted to the real positive axis. In most applications of Eq. 2.70 in this thesis the ω-integration is carried out numerically. Therefore, the present form is not ideal as $\chi(r_1, r_2, \omega)$ has poles close to the real axis, as can be inferred from the spectral representation (Eq. 2.68) which makes numerical integration very tedious. To avoid this, one applies contour integration in combination with the residue theorem. This states that as no poles lie within the upper right complex ω-plane, any contour integral surrounding that area has to yield zero. Thus the contour shown in Fig. 2.1 implies that

$$\int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \chi(r_1, r_2, \omega) = \int_0^\infty \frac{\mathrm{d}\mathrm{i}\omega}{2\pi} \chi(r_1, r_2, \mathrm{i}\omega), \tag{2.71}$$

as the arc connecting real and imaginary ω-infinity does not contribute to the integral as for $\lim_{|\omega| \to \infty} \chi(r_1, r_2, \mathrm{i}\omega) \to 0$. This allows to exchange the frequency integration over real ω by a integration over imaginary ω:

$$E_\mathrm{C} = -\int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \int\int \mathrm{d}r_1 \mathrm{d}r_2 v(r_1, r_2) [\chi_\lambda(r_1, r_2, \mathrm{i}\omega) - \chi_0(r_1, r_2, \mathrm{i}\omega)]. \tag{2.72}$$

Therefore, the pole-shifts in Eq. 2.68 become obsolete and it is possible to omit them in the following. The result in Eq. 2.72 is often referred to as the adiabatic-connection fluctuation-dissipation theorem (ACFDT), which is the basis for the correlation energy expressions derived in the remainder of this chapter.

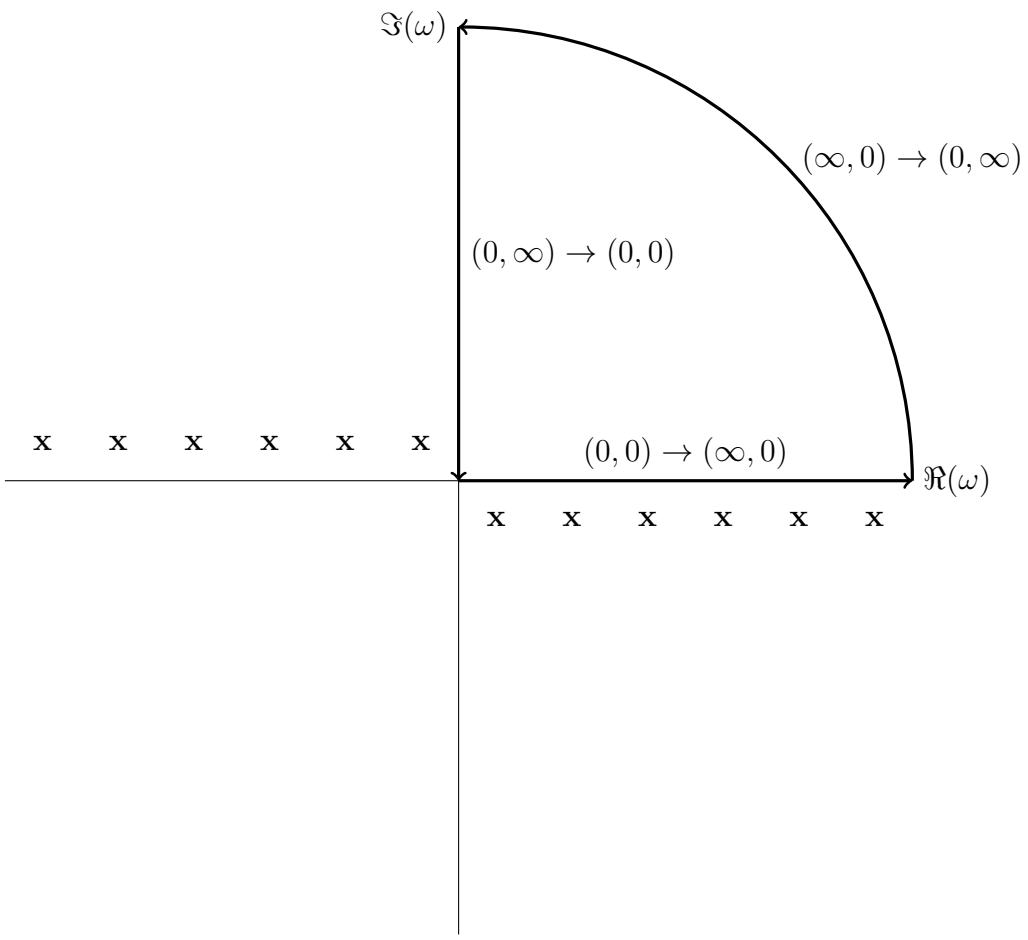


Figure 2.1: View of the complex ω-plane, where the poles of the time-ordered response function $\chi(r_1, r_2, \omega)$ are denoted by $\mathbf{x}$ (compare Eq. 2.68). Furthermore, a contour is shown starting at the origin, going to real infinity, from there to the imaginary axis to imaginary infinity and back to the origin. As there is no pole enclosed by the contour, the contour integral has to evaluate to zero according to the residue theorem.

2.4.2 Strategies to Obtain Approximate Response Functions

What Eq. 2.72 has achieved is replacing the search for the exchange-correlation functional in conventional DFT with the search for the response function along the AC path. For $\lambda = 0$ it is straightforward to find a form that can be evaluated with the one-particle functions and energies of the non-interacting system. In these terms $\chi_0(r_1, r_2, \mathrm{i}\omega)$ is given as

$$\begin{aligned}\chi_0(r_1, r_2, \mathrm{i}\omega) = \sum_{i}^{N_{\mathrm{occ}}} \sum_{a}^{N_{\mathrm{unocc}}} \sum_{\sigma_1,\sigma_2} \Bigg[& \frac{\varphi_i^*(r_1\sigma_1)\varphi_a(r_1\sigma_1)\varphi_i(r_2\sigma_2)\varphi_a^*(r_2\sigma_2)}{\mathrm{i}\omega - (\epsilon_a - \epsilon_i)} \\ & - \frac{\varphi_i^*(r_2\sigma_2)\varphi_a(r_2\sigma_2)\varphi_i(r_1\sigma_1)\varphi_a^*(r_1\sigma_1)}{\mathrm{i}\omega + (\epsilon_a - \epsilon_i)} \Bigg],\end{aligned} \tag{2.73}$$

where the sums run over occupied and unoccupied MOs, respectively. It remains to find expressions for the response function for $\lambda \neq 0$. One popular avenue is to employ TD-DFT linear-response theory. In this field, the retarded response function, defined as

$$\chi_R(r_1 t_1, r_2 t_2) = -\mathrm{i}\theta(t_1 - t_2) \left\langle \Psi_0 \right| \left[\hat{n}(r_1 t_1), \hat{n}(r_2 t_2)\right] \left| \Psi_0 \right\rangle, \tag{2.74}$$

has been the subject of extensive research (see, e.g., Ref. [64] and references therein). One can show that the frequency-domain representation of χ_R is identical to Eq. 2.68 except for the sign of one of the infinitesimal pole shifts. Because the pole shifts are irrelevant for the evaluation of Eq. 2.72, one can use the techniques of TD-DFT to find practical approximations for the interacting response function along the AC path. An important result from TD-DFT is a Dyson-type equation for the response function, which relates the non-interacting response function (Eq. 2.73) to the interacting response function [65]

$$\begin{aligned}\chi_R(r_1, r_2, \omega) = & \chi_{R,0}(r_1, r_2, \omega) \\ & + \int\int \mathrm{d}r_3 \mathrm{d}r_4 \chi_{R,0}(r_1, r_3, \omega)[v(r_3, r_4) \\ & \qquad + f_{\mathrm{xc}}(r_3, r_4, \omega)]\chi_R(r_4, r_2, \omega).\end{aligned} \tag{2.75}$$

Here, f_{xc} is the functional derivative of the frequency dependent exchange-correlation potential with respect to the density. While f_{xc} in Eq. 2.75 is still unknown, one either neglects it, or uses approximations.

An alternative to linear-response TD-DFT is to employ the techniques of MBPT. Here, one goes back to the definition of the time-ordered response function (Eq. 2.60) and realizes that the main ingredient is the expectation value of a time-ordered product of Heisenberg operators with the (λ)-interacting ground state. This is amenable to the adiabatic switching technique [62], which in combination with the Gell-Mann Low theorem [66] and Wick's theorem [67] can be used to evaluate the interacting response function with the help of Feynman diagrams [68, 69]. The simplest building tool in MBPT is the non-interacting Green's function $G_0(r_1 t_1, r_2 t_2)$ which role in low-scaling algorithms for ACFDT and other perturbative approaches is discussed in Section 2.5.2. For a comprehensive account of the abovementioned concepts the reader is referred to Ref. [62]. The simplest approximation to the interacting response function using either of the two techniques is the random-phase-approximation, which is discussed in more detail in the next section.

2.4.3 The Random-Phase-Approximation

As mentioned above, the simplest approximation to an interacting response function is the random-phase-approximation. It is defined by neglecting f_{xc} in Eq. 2.75 altogether, which yields the RPA response function

$$\begin{aligned}\chi^{\mathrm{RPA}}(r_1, r_2, \omega) =& \chi_0(r_1, r_2, \omega) \\ &+ \int\int \mathrm{d}r_3 \mathrm{d}r_4 \chi_0(r_1, r_3, \omega) v(r_3, r_4) \chi^{\mathrm{RPA}}(r_4, r_2, \omega).\end{aligned} \tag{2.76}$$

To evaluate Eq. 2.72, the response function along the AC path is required. Therefore, Eq. 2.75 is generalized to the coupling-strength Hamiltonian and one obtains

$$\chi_\lambda(\omega) = \chi_0(\omega) + \chi_0(\omega)[\lambda v + f_{\mathrm{xc}}^\lambda(\omega)]\chi_\lambda(\omega), \tag{2.77}$$

$$\chi_\lambda^{\mathrm{RPA}}(\omega) = \chi_0(\omega) + \chi_0(\omega)\lambda v \chi_\lambda^{\mathrm{RPA}}(\omega), \tag{2.78}$$

where, from here on, spatial coordinates are omitted, and integration is implicitly expressed via the Einstein sum convention [70]. Plugging Eq. 2.78 into Eq. 2.72 yields the RPA correlation energy

$$E_{\mathrm{C}}^{\mathrm{RPA}} = -\int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{v[\chi_\lambda^{\mathrm{RPA}}(\mathrm{i}\omega) - \chi_0(\mathrm{i}\omega)]\}, \tag{2.79}$$

$$\chi_\lambda^{\mathrm{RPA}}(\mathrm{i}\omega) = (1 - \chi_0(\mathrm{i}\omega)\lambda v)^{-1}\chi_0(\mathrm{i}\omega), \tag{2.80}$$

$$E_{\mathrm{C}}^{\mathrm{RPA}} = -\int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{v[(1 - \chi_0(\mathrm{i}\omega)\lambda v)^{-1}\chi_0(\mathrm{i}\omega) - \chi_0(\mathrm{i}\omega)]\}, \tag{2.81}$$

$$E_{\mathrm{C}}^{\mathrm{RPA}} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\log(1 - \chi_0(\mathrm{i}\omega)v) + \chi_0(\mathrm{i}\omega)v\}, \tag{2.82}$$

where, in Eq. 2.82, the coupling strength integration is carried out analytically. The trace implies a sum over all spatial coordinates. Eq. 2.82 is typically the starting point for most RPA implementations. Since, in quantum chemistry, one generally works with real MOs, it is possible to simplify the expression for the non-interacting response function (Eq. 2.73), in particular for imaginary frequencies:

$$\begin{aligned}\chi_0(r_1, r_2, \mathrm{i}\omega) =& \sum_i^{N_{\mathrm{occ}}} \sum_a^{N_{\mathrm{unocc}}} \sum_{\sigma_1,\sigma_2} \Bigg[\frac{\varphi_i(r_1\sigma_1)\varphi_a(r_1\sigma_1)\varphi_i(r_2\sigma_2)\varphi_a(r_2\sigma_2)}{\mathrm{i}\omega - (\epsilon_a - \epsilon_i)} \\ &- \frac{\varphi_i(r_2\sigma_2)\varphi_a(r_2\sigma_2)\varphi_i(r_1\sigma_1)\varphi_a(r_1\sigma_1)}{\mathrm{i}\omega + (\epsilon_a - \epsilon_i)} \Bigg]\end{aligned} \tag{2.83}$$

$$= \sum_i^{N_{\mathrm{occ}}} \sum_a^{N_{\mathrm{unocc}}} \sum_{\sigma_1,\sigma_2} \frac{-2(\epsilon_a - \epsilon_i)\varphi_i(r_1\sigma_1)\varphi_a(r_1\sigma_1)\varphi_i(r_2\sigma_2)\varphi_a(r_2\sigma_2)}{\omega^2 + (\epsilon_a - \epsilon_i)^2}. \tag{2.84}$$

Furthermore, one chooses to work in the MO basis rather than in the real-space basis. To transform Eq. 2.82 into the MO basis, the quantities

$$V_{ia,jb} = (ia|jb) = \int\int \mathrm{d}x_1 \mathrm{d}x_2 \frac{\varphi_i(x_1)\varphi_a(x_1)\varphi_j(x_2)\varphi_b(x_2)}{r_{12}}, \tag{2.85}$$

$$\Pi_{0_{ia,jb}}(\mathrm{i}\omega) = \delta_{ij}\delta_{ab}\frac{-2(\epsilon_a - \epsilon_i)}{\omega^2 + (\epsilon_a - \epsilon_i)^2}, \tag{2.86}$$

are introduced, which will be referred to as the Hartree-kernel (Eq. 2.85) and the non-interacting polarization propagator (Eq. 2.86), and which are the particle-hole basis representation of the Coulomb operator and the non-interacting response function, respectively. Using the invariance of the trace under cyclic permutations allows to introduce $\mathbf{V}$ and $\mathbf{\Pi}_0(\mathrm{i}\omega)$ into Eq. 2.82

$$E_\mathrm{C}^\mathrm{RPA} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\log(1 - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}) + \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\} \tag{2.87}$$

where the trace now implies summation over particle-hole indices. Eq. 2.87 can now be used to evaluate the RPA correlation energy with MOs from a preceding self-consistent KS calculation. With a formal $\mathcal{O}(N^6)$ scaling behavior and the requirement to perform the numerical frequency integration, Eq. 2.87 is not yet ideal. Low-scaling RPA approaches will be introduced in Section 2.5. It is possible to perform the frequency integration in Eq. 2.87 analytically, which leads to the so called plasmon formula [22, 49]. While this circumvents the numerical frequency integration, it still shows a prohibitive $\mathcal{O}(N^6)$ scaling behavior. Furthermore, it has been shown [71] that direct ring Coupled Cluster is equivalent to the RPA method introduced here. As the low-scaling methods presented in this work relate to the dielectric matrix formalism (Eq. 2.87) further discussion of the plasmon formula and the CC variant of RPA are omitted.

2.4.4 Beyond RPA: Inclusion of Exchange Effects

While the RPA has a variety of features that make it an appealing method, such as the good description of dispersion effects [14, 15], or the ability to treat vanishing gap sytems, it also has a variety of shortcomings. These include unsatisfactory performance for short-range correlation effects and bad performance for non-isogyric processes [72–75]. These failures can be traced back to the self-interaction in the RPA correlation energy due to the complete neglect of exchange effects since f_xc is set to zero. The simplest example showing this deficiency is that the RPA yields non-zero correlation energies for one-electron systems [23], which will be discussed further below. Therefore, there is a need to include exchange into RPA correlation energies to remedy these issues and further increase the accuracy of RPA correlation energies.

Derivation of Beyond RPA Methods

While one could also work with Eq. 2.75 to derive more accurate response functions, all RPA with exchange schemes that are used in this work were derived from the analogous formula for the λ-interacting polarization propagator $\mathbf{\Pi}_\lambda(\mathrm{i}\omega)$

$$\mathbf{\Pi}_\lambda(\mathrm{i}\omega) = \mathbf{\Pi}_0(\mathrm{i}\omega) + \mathbf{\Pi}_0(\mathrm{i}\omega)[\lambda\mathbf{V} + \mathbf{F}_\mathrm{xc}^\lambda(\mathrm{i}\omega)]\mathbf{\Pi}_\lambda(\mathrm{i}\omega). \tag{2.88}$$

Eq. 2.88 is also known as Bethe-Salpeter equation [76]. Here, the RPA corresponds to setting $\mathbf{F}_\mathrm{xc}^\lambda(\mathrm{i}\omega) = \mathbf{0}$. The inclusion of exchange-effects is performed by finding an approximate form for the exchange-correlation kernel $\mathbf{F}_\mathrm{xc}^\lambda$, which reduces or removes the self-interaction in the RPA. One approach suggested by Mussard et al. [22] is to use a static, approximate exchange kernel

$$K_{ia,jb} = (ib|ja), \tag{2.89}$$

to obtain the RPA with exchange (RPAx) polarization propagator

$$\mathbf{\Pi}_\lambda^{\mathrm{RPAx}}(\mathrm{i}\omega) = \mathbf{\Pi}_0(\mathrm{i}\omega) + \mathbf{\Pi}_0(\mathrm{i}\omega)[\lambda(\mathbf{V}-\mathbf{K})]\mathbf{\Pi}_\lambda^{\mathrm{RPAx}}(\mathrm{i}\omega). \tag{2.90}$$

This can be used together with the particle-hole basis variant of the ACFDT formula for the correlation energy (Eq. 2.72)

$$E_\mathrm{C} = -\int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\mathbf{\Pi}_\lambda(\mathrm{i}\omega)\mathbf{V} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\}, \tag{2.91}$$

to obtain the RPAx correlation energy

$$E_\mathrm{C}^{\mathrm{RPAx}} = -\int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\mathbf{\Pi}_\lambda^{\mathrm{RPAx}}(\mathrm{i}\omega)\mathbf{V} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\}. \tag{2.92}$$

After analytical coupling strength integration one obtains

$$E_\mathrm{C}^{\mathrm{RPAx}} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\log(1 - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{W})\mathbf{W}^{-1}\mathbf{V} + \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\}, \tag{2.93}$$

with $\mathbf{W} = \mathbf{V} - \mathbf{K}$. The RPAx correlation energy is free of one-electron self-interaction and describes the dissocation of H_2 correctly (vide infra). Expanding the logarithm to second order yields the MP2 correlation energy expressed in terms of an imaginary frequency integral

$$E_{\mathrm{C},(II)}^{\mathrm{RPAx}} = -\frac{1}{2}\int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{W}\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\} = E_\mathrm{C}^{\mathrm{MP2}}. \tag{2.94}$$

More correlation effects are described by the infinite sum implied by the matrix logarithm. There is, however, a significant problem with this approach: Compared to the argument of the logarithm in the RPA correlation energy expression, $\mathbf{1} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}$ (Eq. 2.87), whose eigenvalue spectrum is bounded from below by 0, the RPAx argument $\mathbf{1} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{W}$ is indefinite. This can lead to cases where one cannot evaluate the argument of the logarithm due to the occurrence of negative eigenvalues. This phenomenon is known as triplet instability and is also known in the ring CC variant of RPA methods [19]. Other approaches using the exact exchange kernel from TD-DFT instead of $\mathbf{K}$ suffer from the same problem [77, 78].

To circumvent this problem one could try to include additional terms into $\mathbf{F}_\mathrm{xc}$. As the RPAx equations are, however, quite computationally involved and adding more terms leads to further complications, other approximate RPA with exchange schemes with similar computational complexity are desirable.

One method proposed by Furche and coworkers [73] consists of approximating the RPAx polarization propagator (Eq. 2.90) by a series expansion in the RPA polarization propagator

$$\begin{aligned}
\mathbf{\Pi}_\lambda^{\mathrm{RPA}}(\mathrm{i}\omega) &= (\mathbf{\Pi}_0(\mathrm{i}\omega)^{-1} - \lambda\mathbf{V})^{-1}, & (2.95)\\
\mathbf{\Pi}_\lambda^{\mathrm{RPAx}}(\mathrm{i}\omega) &= (\mathbf{\Pi}_0(\mathrm{i}\omega)^{-1} - \lambda\mathbf{V} + \lambda\mathbf{K})^{-1} = (\mathbf{\Pi}_\lambda^{\mathrm{RPA}^{-1}}(\mathrm{i}\omega) + \lambda\mathbf{K})^{-1}, & (2.96)\\
&= \mathbf{\Pi}_\lambda^{\mathrm{RPA}}(\mathrm{i}\omega) - \mathbf{\Pi}_\lambda^{\mathrm{RPA}}(\mathrm{i}\omega)\lambda\mathbf{K}\mathbf{\Pi}_\lambda^{\mathrm{RPA}}(\mathrm{i}\omega) + \ldots. & (2.97)
\end{aligned}$$

Including only the term in first order of $\mathbf{K}$ additionally defines the approximate-exchange-kernel (AXK) correction to RPA correlation energies. After coupling-strength integration one obtains the following expression for the AXK correction

$$\Delta E_\mathrm{C}^{\mathrm{AXK}} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\log(1 - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V})\mathbf{V}^{-1}\mathbf{K} + (1 - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V})^{-1}\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\}. \tag{2.98}$$

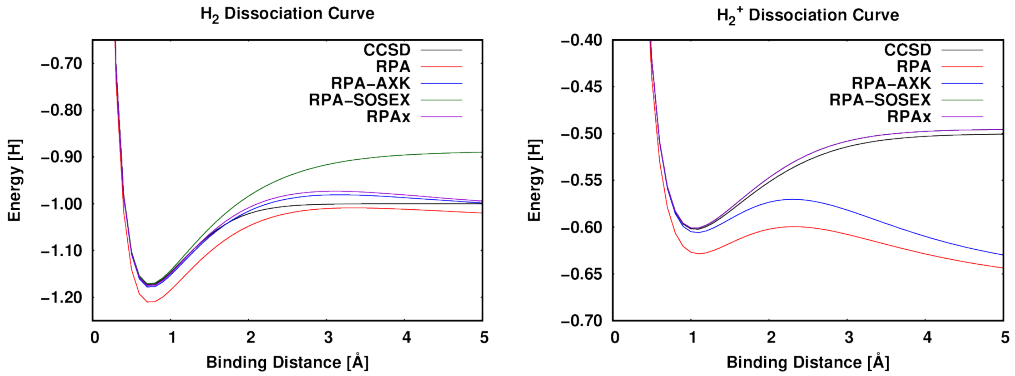


Figure 2.2: H_2- and H_2^+-dissociation curves for RPA, RPAx, RPA-SOSEX and RPA-AXK calculations using a PBE reference and the def2-QZVP basis set. The CCSD curve represents the exact curve.

This circumvents the problem of the RPAx energy expression (Eq. 2.92). Before the discussion of the strengths- and weaknesses of the RPA-AXK approach the last RPA with exchange scheme that was used in this work is introduced. The second-order screened exchange (SOSEX) correction to RPA was first introduced in the CC variant of RPA [20, 21]. In the ACFDT framework it consists of using the RPA polarization propagator and exchanging the Hartree-kernel $\mathbf{V}$ with the approximate exchange kernel $\mathbf{K}$ in Eq. 2.91

$$\Delta E_{\mathrm{C}}^{\mathrm{SOSEX}} = \int_0^1 \mathrm{d}\lambda \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\mathbf{\Pi}_\lambda^{\mathrm{RPA}}(\mathrm{i}\omega)\mathbf{K} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\}, \tag{2.99}$$

$$\Delta E_{\mathrm{C}}^{\mathrm{SOSEX}} = -\int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\log(1 - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V})\mathbf{V}^{-1}\mathbf{K} + \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\}, \tag{2.100}$$

where Eq. 2.100 is obtained after performing the coupling strength integration. Since the RPA polarization propagator is used and exchange is introduced by modification of the electron-electron interaction operator, RPA-SOSEX also does not suffer from the triplet instability. From an algebraic standpoint, the series expansion of both the RPA-AXK and the RPA-SOSEX energy expression also yield MP2 at second order and they only start to deviate from third order [73].

Comparison of Different RPA with Exchange Methods

To examine the strengths and weaknesses of the abovementioned methods, the model systems H_2 and H_2^+ serve as good test cases, because in the dissociation limit they exhibit static correlation or self-interaction errors, respectively. A large variety of electronic structure methods fail to describe both of them correctly. As can be seen in Figure 2.2, also RPA and beyond RPA methods struggle to obtain the correct dissociation limits. In fact, only the RPAx method delivers a qualitatively correct description of both systems. It is one-electron self-interaction free and describes the static correlation in dissociated H_2 correctly. Unfortunately, as already mentioned above, due to the triplet instability, the RPAx method is not universally applicable to all chemical systems, failing, e.g., for O_2. RPA and RPA-AXK mimic the static correlation present in dissociated H_2 well. For the one-electron system of dissoci-

ated H_2^+ both methods, however, fail due to self-interaction errors, which are less pronounced in RPA-AXK. RPA-SOSEX shows the exact opposite behavior being exact for H_2^+, but failing to describe the static correlation in the dissocation limit of H_2. The approximate RPA with exchange methods RPA-AXK and RPA-SOSEX improve significantly over RPA in the equilibrium distance of both molecules.
The failure of RPA-SOSEX for stretched H_2 inspired the development of a new RPA with exchange scheme in **Publication IV**. Realizing that SOSEX accurately corrects the short-range behavior of RPA, but leads to a worse description in the long-range regime, has motivated a short-range SOSEX correction. The novel short-range SOSEX scheme allows to combine the benefits of RPA and RPA-SOSEX, resulting in a a better performance across a variety of problems including static correlation and self-interaction error. This includes barrier heights, electron affinities, ionization potentials and electron deficient dimers.
While including exchange effects in RPA correlation energies increases the accuracy for a variety of properties, it also increases the computational cost significantly. Before this thesis the most efficient ACFDT RPA with exchange schemes showed an $\mathcal{O}(N^5)$ scaling behavior, limiting the application of these methods to small systems. In **Publication V** linear- and low scaling methods for all of the abovementioned RPA with exchange methods were presented. This allows to further explore and apply the benefits of RPA with exchange methods to significantly larger systems than possible before. The newly developed methods enabled to show that RPA-SOSEX is more accurate than plain RPA and other electron correlation methods such as MP2 for the large dispersion dominated systems in the L7 test set [24].

2.5 Numerical Approximations for Low-Scaling Quantum Chemical Methods

This section introduces the techniques developed and employed in this thesis to make the electron correlation methods presented in Section 2.4 applicable to large chemical systems. While the physical approximations introduced in the traditional derivation of the RPA and beyond RPA schemes have circumvented the exponential scaling of FCI, they still scale as $\mathcal{O}(N^6)$ with the system size in the canonical formalism (see Eqs. 2.87, 2.92, 2.98 and 2.100). While this allows to treat small to medium sized molecular systems, it is still prohibitively expensive for large molecules. As this is also the case for other electron correlation methods such as CC and MP2, a large arsenal of techniques have been developed to reduce the scaling behavior of these methods without changing the underlying physical model. The popular field of local correlation methods [79] relies on techniques to localize the MOs through unitary transformations. This localization then potentially leads to a greatly reduced number of significant interactions, resulting in low scaling methods. While there has been work on RPA techniques using localized MOs [80, 81], this will not be the focus here. Instead, techniques to reduce the scaling behavior are introduced that rely on the sparsity of AOs and one-particle density matrices. Furthermore, a quick account on the resolution-of-the-identity (RI) is given, which has widespread use in efficient quantum chemical methods, as it allows for a compact encapsulation of quantities such as the non-interacting response function in auxiliary gaussian basis sets.

2.5.1 Resolution-of-the-Identity

The resolution-of-the-identity [82–85] is a widespread concept in quantum chemistry and physics, which relies on the insertion of an identity spanned by an auxiliary basis set

$$\hat{1} = |\tilde{P})(\tilde{P}|. \tag{2.101}$$

For Eq. 2.101 to hold, $\{\tilde{P}\}$ has to be a complete orthogonal set. While completeness can only ever be approximated with finite basis sets, the orthogonality has also to be taken care of, as the gaussian basis sets typically applied for the RI are not orthogonal. As the RI is mainly used to represent charge distributions spanned by AO basis function pairs, one of the first strategies to obtain practical equations for Eq. 2.101 used the following fitting strategy [83]

$$\frac{\partial}{\partial K_P^{\mu\nu}}(\mu\nu - \tilde{\mu\nu}|m_{12}|\mu\nu - \tilde{\mu\nu}) = 0, \tag{2.102}$$

$$|\tilde{\mu\nu}) = \sum_P^{N_{\text{aux}}} K_P^{\mu\nu}|\chi_P), \tag{2.103}$$

where $|\mu\nu)$ is a product of the primary AO basis functions, and the integral in Eq. 2.102 denotes a two-electron integral in the Mulliken notation with the two point function m_{12}. Auxiliary basis functions are denoted by $P, Q, \cdots$. The coefficents $K_P^{\mu\nu}$ were then obtained by use of Eq. 2.102, which minimises the self-repulsion of the difference charge-density with respect to the operator m_{12}. Historically, m_{12} was chosen as the overlap metric ($\delta_{r_{12}}$) or the Coulomb metric ($\frac{1}{r_{12}}$). Due to the procedure described above, the RI is also widely known as density fitting [86]. The most general formula for Eq. 2.101 in an approximate, non-orthogonal basis set is presumably

$$\hat{1} \approx |P)(P|m_{12}|Q)^{-1}(Q|m_{12}, \tag{2.104}$$

which can be derived either from Eq. 2.102 or purely from an RI perspective [87]. This can now be used to, e.g., factorize a four-center, two-electron integral in two- and three-center quantities

$$(\mu\nu|\lambda\sigma) \approx (\mu\nu|m_{12}|P)(P|m_{12}|Q)^{-1}(Q|\frac{1}{r_{12}}|R)(R|m_{12}|S)^{-1}(S|m_{12}|\lambda\sigma), \tag{2.105}$$

which is employed in different areas of quantum chemistry from self-consistent field to correlated calculations.
In the context of RPA calculations the RI was introduced by Furche and coworkers [13] to obtain an $\mathcal{O}(N^4)$ scaling method to calculate RPA correlation energies. This is in contrast to the naive $\mathcal{O}(N^6)$ scaling method when directly evaluating e.g. Eq. 2.87. While not the original derivation of the RI-RPA method, the quickest way to obtain the RI-RPA expression is to insert Eq. 2.104 into Eq. 2.82 and use the invariance of the trace in combination with the series expansion of the logarithm to obtain a formula in terms of the auxiliary space representations of the Coulomb

operator and the non-interacting response function

$$E_{\mathrm{C}}^{\mathrm{RPA}} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\log(1 - \chi_0(\mathrm{i}\omega)\hat{1}v\hat{1}) + \chi_0(\mathrm{i}\omega)\hat{1}v\hat{1}\}, \tag{2.106}$$

$$E_{\mathrm{C}}^{\mathrm{RPA}} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\log(1 - \chi_0(\mathrm{i}\omega)\mathbf{C}) + \chi_0(\mathrm{i}\omega)\mathbf{C}\}, \tag{2.107}$$

$$\chi_0(\mathrm{i}\omega)_{PQ} = (P|m_{12}|ia)\Pi_0(\mathrm{i}\omega)_{ia,jb}(jb|m_{12}|Q) \tag{2.108}$$

$$C_{PQ} = (P|m_{12}|R)^{-1}(R|S)(S|m_{12}|Q)^{-1}. \tag{2.109}$$

The benefit of Eq. 2.107 over Eq. 2.87 is that the $\mathcal{O}(N_{\mathrm{occ}}^3 N_{\mathrm{unocc}}^3)$ matrix multiplications are avoided and the most expensive step, namely the calculation of $\chi_0(\mathrm{i}\omega)$, scales only as $\mathcal{O}(N_{\mathrm{aux}}^2 N_{\mathrm{occ}} N_{\mathrm{unocc}})$. As the dimension of the auxiliary basis set N_{aux} scales linearly with the system size, the scaling is reduced from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^4)$. Furche and coworkers [13] employed the Coulomb metric for the RI decomposition. In contrast, the linear scaling AO based formulations of the RPA use either the overlap [16] or the attenuated-Coulomb metric $m_{12} = \frac{\mathrm{erfc}(wr_{12})}{r_{12}}$ [17], which ensure that only a linear number of elements of the three-center integral tensors $\mathbf{B}$ are significant in a local basis such as the AO basis

$$B_{\mu\nu}^P = (P|m_{12}|\mu\nu). \tag{2.110}$$

While the overlap metric is more local than the attenuated Coulomb metric, it is much less accurate [17]. This shows that by the introduction of the RI, the formal scaling behavior of an electronic structure method can be reduced independently of the properties of the molecular system. The next section illustrates how to further optimize the scaling behavior by exploiting locality in the electronic structure.

2.5.2 Atomic Orbital Formalism

Locality of Atomic Orbital Quantitites: The Example of Exchange

Using AOs has been a successful strategy to obtain efficient electronic structure methods for large molecules. A good example to show the benefits of using AOs is the formation of the exchange matrix in SCF calculations. In terms of AOs the working equation is

$$K_{\mu\nu} = \sum_{\lambda,\sigma}^{N_{\mathrm{bas}}} P_{\lambda\sigma}(\mu\lambda|\nu\sigma), \tag{2.111}$$

$$P_{\mu\nu} = \sum_{i}^{N_{\mathrm{occ}}} C_{\mu i} C_{\nu i}, \tag{2.112}$$

where $\mathbf{P}$ is the one-particle density matrix and C contains the MO coefficients. While the naive calculation of $\mathbf{K}$ scales formally as $\mathcal{O}(N_{\mathrm{bas}}^4)$ with the basis set size N_{bas}, this can be reduced to linear scaling using only two crucial observations. First, in the asymptotic limit of large molecules there is only a linear number of significant charge distributions $|\mu\nu)$ when gaussian basis functions are employed, as the product of two gaussians decays exponentially with the distance between the center of the basis functions. This fact can be used to select a quadratic number of significant integrals through the use of Schwarz's inequality [88]

$$(\mu\nu|\lambda\sigma) \leq \sqrt{(\mu\nu|\mu\nu)}\sqrt{(\lambda\sigma|\lambda\sigma)}. \tag{2.113}$$

Second, **P** contains only a linear number of significant elements for extended systems with non-vanishing HOMO-LUMO gap [89–91]. This allows to reduce the number of significant contractions to linear by using the following screening criterium

$$P_{\lambda\sigma}(\mu\lambda|\nu\sigma) \leq |P_{\lambda\sigma}|\sqrt{(\mu\lambda|\mu\lambda)}\sqrt{(\nu\sigma|\nu\sigma)} \tag{2.114}$$

where the one particle density matrix renders only a constant numbers of Kets significant for each Bra. A clever hierarchy of screening loops has therefore allowed to reduce the scaling of the exchange matrix formation to linear for large, gapped systems [92–94]. **Publication I** deals with further optimization of these strategies for the local attenuated Coulomb operator in short-range hybrid DFT calculations through the use of adapted integral screening criteria and screening schemes for CPU and GPU architectures.

Atomic Orbitals in Electron Correlation: The Example of MP2

As shown above, AOs allow for the development of low-scaling electronic structure methods. For correlated methods, devising AO schemes needs further attention, since usually the working equations are given in terms of non-local canonical orbitals. The textbook example is the MP2 energy expression

$$E_{\mathrm{C}}^{\mathrm{MP2}} = -\sum_{i,j}^{N_{\mathrm{occ}}}\sum_{a,b}^{N_{\mathrm{unocc}}} \frac{(ia|jb)[2(ia|jb)-(ib|ja)]}{\epsilon_a+\epsilon_b-\epsilon_i-\epsilon_j}. \tag{2.115}$$

From now on, occupied orbitals are labelled as $i, j, \cdots$ and virtual orbitals as $a, b, \cdots$. Here, the introduction of AOs is hindered by the denominator which couples four MO indices. This was circumvented by the Laplace transformation [95]

$$\frac{1}{\epsilon_a+\epsilon_b-\epsilon_i-\epsilon_j} = \int_0^\infty \mathrm{d}\tau e^{-\epsilon_a\tau}e^{-\epsilon_b\tau}e^{\epsilon_i\tau}e^{\epsilon_j\tau}, \tag{2.116}$$

which allowed to formulate Eq. 2.115 in the AO basis [96, 97],

$$\begin{aligned} E_{\mathrm{C}}^{\mathrm{MP2}} = \int_0^\tau \mathrm{d}\tau \sum_{\mu,\nu,\lambda,\sigma}^{N_{\mathrm{bas}}}\sum_{\mu',\nu',\lambda',\sigma'}^{N_{\mathrm{bas}}} & \underline{P}^\tau_{\mu\mu'}\underline{P}^\tau_{\lambda\lambda'}\overline{P}^\tau_{\nu\nu'}\overline{P}^\tau_{\sigma\sigma'}(\mu\nu|\lambda\sigma) \\ & [2(\mu'\nu'|\lambda'\sigma')-(\mu'\sigma'|\lambda'\nu')]. \end{aligned} \tag{2.117}$$

The occupied and virtual pseudo-densities are defined as

$$\underline{P}^\tau_{\mu\nu} = \sum_i^{N_{\mathrm{occ}}} C_{\mu i}e^{\epsilon_i\tau}C_{\nu i}, \tag{2.118}$$

$$\overline{P}^\tau_{\mu\nu} = \sum_a^{N_{\mathrm{unocc}}} C_{\mu a}e^{-\epsilon_a\tau}C_{\nu a}. \tag{2.119}$$

The occupied and virtual pseudo-densities (Eqs. 2.118, 2.119) are similarly local as the occupied and virtual one-particle density matrices (Eqs. 2.112, 2.140). Eq. 2.117 in combination with schemes exploiting the locality of the AO quantities allowed for a variety of low-scaling MP2 formulations (see, e.g., Refs. [98–102]).

Atomic Orbitals in ACFDT Methods: Imaginary Time Represention of the Non-Interacting Polarizability

The quantity which needs to be decoupled in RPA correlation energy calculations is the non-interacting polarization propagator (Eq. 2.86) or equivalently the non-interacting response function (Eq. 2.73). In the linear-scaling AO-RPA method, Schurkus and Ochsenfeld [16] devised a contracted double Laplace transform that lead to three integral transforms which allow to decouple the MO indices. This approach required two formulations of the integral transforms for numerical stability of the quadrature. **Publication II** uses the equivalence of one of the integral transforms with the Fourier transform of the non-interacting polarization propagator from the imaginary time domain to the imaginary frequency domain. In combination with an improved customized quadrature scheme [103] for the imaginary frequency integral and the Fourier transform, a numerically significantly more accurate and faster RPA method was developed. In particular for the non-interacting response function, this transformation is given as [104, 105]

$$\chi_0(r_1, r_2, \mathrm{i}\omega) = \int_{-\infty}^{\infty} \mathrm{d}\tau e^{\mathrm{i}\omega\tau} \chi_0(r_1, r_2, \mathrm{i}\tau) = \int_{-\infty}^{\infty} \mathrm{d}\tau \cos(\omega\tau) \chi_0(r_1, r_2, \mathrm{i}\tau), \tag{2.120}$$

where the Fourier transform reduces to a cosine transform, as the non-interacting response function is symmetric in $\mathrm{i}\omega$ (compare Eq. 2.84). One important result from MBPT is that the non-interacting response function in the imaginary time domain is given as

$$\chi_0(r_1, r_2, \mathrm{i}\tau) = G_0(r_1, r_2, \mathrm{i}\tau) G_0(r_2, r_1, -\mathrm{i}\tau), \tag{2.121}$$

with the non-interacting Green's function:

$$G_0(r_1, r_2, \mathrm{i}\tau) = \Theta(-\mathrm{i}\tau) \underline{G}_0(r_1, r_2, \mathrm{i}\tau) + \Theta(\mathrm{i}\tau) \overline{G}_0(r_1, r_2, \mathrm{i}\tau), \tag{2.122}$$

$$\underline{G}_0(r_1, r_2, \mathrm{i}\tau) = \sum_{i}^{\mathrm{occ}} \varphi_i(r_1) \varphi_i^*(r_2) e^{-\varepsilon_i \tau} = \sum_{\mu,\nu}^{N_{\mathrm{bas}}} \phi_\mu(r_1) \phi_\nu(r_2) \underline{G}_{0_{\mu\nu}}(\mathrm{i}\tau), \tag{2.123}$$

$$\overline{G}_0(r_1, r_2, \mathrm{i}\tau) = -\sum_{a}^{\mathrm{unocc}} \varphi_a(r_1) \varphi_a^*(r_2) e^{-\varepsilon_a \tau} = -\sum_{\mu,\nu}^{N_{\mathrm{bas}}} \phi_\mu(r_1) \phi_\nu(r_2) \overline{G}_{0_{\mu\nu}}(\mathrm{i}\tau). \tag{2.124}$$

Here, τ is given as $t_1 - t_2$. The atomic orbital Green's functions $\underline{\mathbf{G}}_0(\mathrm{i}\tau)$ and $\overline{\mathbf{G}}_0(\mathrm{i}\tau)$ are equivalent to the occupied (Eq. 2.118) and unoccupied pseudodensities (Eq. 2.119). Eq. 2.121 also translates to the auxiliary basis representation of the non-interacting response function used in RI-RPA

$$\begin{aligned} \chi_0(\mathrm{i}\omega)_{PQ} &= \int_{-\infty}^{\infty} \mathrm{d}\tau \cos(\omega\tau) \chi_0(\mathrm{i}\omega)_{PQ}, \\ &= \int_{-\infty}^{\infty} \mathrm{d}\tau \cos(\omega\tau) B^P_{\mu\nu} G_{\mu\lambda}(-\mathrm{i}\tau) G_{\nu\sigma}(\mathrm{i}\tau) B^Q_{\lambda\sigma}, \end{aligned} \tag{2.125}$$

which yields exactly the time-determining step in the linear-scaling RPA method [16]. This also gives a new perspective on the Laplace transform in MP2 theory. Starting from the spectral representation introduced in Eq. 2.94 in real-space without using the ω-symmetry explicitly

$$E_{\mathrm{C}}^{\mathrm{MP2}} = -\frac{1}{4} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\{\chi_0(\mathrm{i}\omega) w \chi_0(\mathrm{i}\omega) v\}, \tag{2.126}$$

where $w(r_1, r_2) = v(r_1, r_2) - k(r_1, r_2)$ and $k(r_1, r_2) = \hat{P}/r_{12}$, with the same permutation operator as in the case of HF exchange and inserting the Fourier transform for the frequency dependent response function twice yields

$$\begin{aligned} E_{\mathrm{C}}^{\mathrm{MP2}} &= -\frac{1}{4}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{d}\tau\mathrm{d}\tau' e^{\mathrm{i}\omega(\tau+\tau')}\mathrm{Tr}\{\chi_0(\mathrm{i}\tau)w\chi_0(\mathrm{i}\tau')v\} \\ &= -\frac{1}{4}\int_{-\infty}^{\infty}\mathrm{d}\tau\mathrm{Tr}\{\chi_0(\mathrm{i}\tau)w\chi_0(\mathrm{i}\tau)v\}, \end{aligned} \tag{2.127}$$

where the identity $\int \mathrm{d}\omega e^{\mathrm{i}\omega(\tau+\tau')} = 2\pi\delta_{\tau\tau'}$ was used [103]. Using the invariance of the trace and inserting the definition of the non-interacting response function in the imaginary time domain (Eq. 2.121) allows to transform Eq. 2.127 to Eq. 2.117. As the non-interacting Green's functions in Eq. 2.125 and the three-center integral tensor $\mathbf{B}$ become sparse quantities for large molecules, an asymptotic $\mathcal{O}(\mathrm{N})$ scaling RPA method has been achieved [16] and further optimized by Luenser et al. [17] as well as in **Publication II**.

To summarize this section, a strategy was described that allows electronic structure methods that depend on the non-interacting response function to be formulated in AOs through the imaginary time representation of χ_0 which allows for low-scaling methods, if the locality of the resulting quantities is exploited. Also, it was shown how the Laplace transform of MP2 theory fits in this framework.

For practical implementation in Gaussian basis sets, the RI has proved very useful to represent real-space quantities in a compact auxiliary basis. Electron correlation methods that solely depend on the non-interacting response function employ Eq. 2.108 which can be calculated in a linear-scaling manner using Eq. 2.125 with a local-metric [16]. In theories that include exchange, one additionally has to encapsulate the exchange effects which translates into finding a auxiliary basis representation of

$$Y(r_1, r_2, \mathrm{i}\omega) = \chi_0(r_1, r_3, \mathrm{i}\omega)k(r_3, r_4)\chi_0(r_4, r_2, \mathrm{i}\omega). \tag{2.128}$$

This is given as [22]

$$Y_{PQ}(\mathrm{i}\omega) = \sum_{i,j}^{N_{\mathrm{occ}}}\sum_{a,b}^{N_{\mathrm{unocc}}} B_{ia}^{P}\Pi_{0_{ia,ia}}(\mathrm{i}\omega)(ib|ja)\Pi_{0_{jb,jb}}(\mathrm{i}\omega)B_{jb}^{Q}. \tag{2.129}$$

Forming $\mathbf{Y}(\mathrm{i}\omega)$ is the time-determining step in RPA with exchange schemes employing the RI [22]. The formal scaling behavior of this step is $\mathcal{O}(N^5)$, i.e., one power worse than plain RI-RPA. **Publication V** introduces methods to calculate $\mathbf{Y}(\mathrm{i}\omega)$ for large systems in a linear- and low-scaling manner. These methods also rely on imaginary time representations of the response function in combination with exploiting the locality of exchange type contractions similar to Eq. 2.111. These developments allow to calculate RPA with exchange correlation energies for systems with up to 500 atoms, significantly extending the scope of these methods, which was limited to small systems before due to the $\mathcal{O}(N^5)$ scaling behavior.

2.5.3 Cholesky Orbitals Basis for Correlated Electronic Structure Theory

There are numerous examples that show that through the use of AOs the scaling behavior of electronic structure methods can be reduced significantly. One problem

that arises, however, with the use of AO methods is the redundancy present in typical AO basis set. In the absence of sparsity, this leads to a significant computational overhead as, e.g., a sum over occupied orbitals will run over the dimension of the basis set instead. For correlated electronic structure theory, where large basis sets are required, this can be a factor of 10 times more. To tackle this problem, rank reduction through pivoted Cholesky decomposition (CD) [17, 106–110] has proved very useful. Given a real, symmetric, positive semi-definite matrix $\mathbf{A}$ its CD is given as

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T \tag{2.130}$$

where $\mathbf{L}$ denotes the matrix of Cholesky vectors with the same number of rows as the original matrix, but only $rank(\mathbf{A})$ non-zero columns. In electronic structure theory this can be used to decompose, e.g., the occupied one particle density matrix $\mathbf{P}$, which only has rank N_{occ}, as can be seen from its spectral form (Eq. 2.112). Therefore, with an adequate rearrangement of the summations, the canonical formal scaling behavior can be reestablished while using CD quantities. The striking feature of the Cholesky vectors is that they retain the sparsity of AO quantities without increasing the formal scaling behavior of the method. This scheme was extended to correlated electronic structure calculations at the MP2 [98] and RPA [17] level, where instead the occupied and unoccupied pseudo-densities Eqs. 2.118, 2.119 (or the equivalent AO Green's functions) have been decomposed. While this allows for low-scaling algorithms, multiple pseudo-densitities have to be decomposed and, e.g., in the CDD-RI-RPA method [17], the three center integrals are transformed with the Cholesky vectors from every occupied pseudo-density. For computational efficiency this requires to keep the three-center integrals in the AO basis in computer memory, which results in a bottleneck in the calculation of large systems. In **Publication II**, a scheme was developed that allows to circumvent the abovementioned drawbacks. Comparing the CD of the one particle density matrix

$$\mathbf{P} = \mathbf{L}\mathbf{L}^T \tag{2.131}$$

with its spectral decomposition

$$\mathbf{P} = \mathbf{C}\mathbf{N}\mathbf{C}^T, \tag{2.132}$$

where $\mathbf{C}$ contains all MOs and $\mathbf{N}$ is the occupation number matrix, allows to deduce the following form of the Cholesky vectors

$$\mathbf{L} = \mathbf{C}\mathbf{N}^{1/2}\mathbf{U}, \tag{2.133}$$

where $\mathbf{U}$ is a unitary matrix. As $\mathbf{N}$ contains only N_{occ} ones on its diagonal, only the occupied MOs and the N_{occ} orthonormal vectors of $\mathbf{U}$ are used. Solving for these vectors $\mathbf{U}^{\text{occ}}$

$$\mathbf{U}^{\text{occ}} = \mathbf{C}_{\text{occ}}\mathbf{S}\mathbf{L} \tag{2.134}$$

gives a $N_{\text{occ}} \times N_{\text{occ}}$ unitary matrix that can be used to transform occupied orbitals to the occupied Cholesky orbital basis (compare also Ref. [111]). Here $\mathbf{S}$ is the AO overlap matrix. Performing this transformation for the occupied Green's function from the canonical into the local basis of Cholesky orbitals (denoted by $\underline{i}, \underline{j}, \cdots$) yields

$$\underline{G}_0(\mathrm{i}\tau)_{\underline{i}\underline{j}} = U^{\text{occ}^\dagger}_{\underline{i}i}\, \underline{G}_0(\mathrm{i}\tau)_{ij} U^{\text{occ}}_{j\underline{j}}. \tag{2.135}$$

The negative imaginary time Green's function in the Cholesky orbital basis $\underline{G}_0(\mathrm{i}\tau)_{ij}$ is a key ingredient in **Publications II, III, V** as it offers a compact and local representation. Revisiting for example the time determining step in RI-RPA, namely the formation of the non-interacting response function in imaginary time, starting from the canonical MO basis

$$\chi_{0_{PQ}}(\mathrm{i}\tau) = B^P_{ia}\underline{G}_0(\mathrm{i}\tau)_{ij}\overline{G}_0(\mathrm{i}\tau)_{ab}B^Q_{jb} \tag{2.136}$$

yields the formulation used in **Publication II** after inserting $\mathbf{1} = \mathbf{U}^{\mathrm{occ}}\mathbf{U}^{occ,\dagger}$ twice

$$\begin{aligned}\chi_{0_{PQ}}(\mathrm{i}\tau) =& B^P_{ia}(U^{\mathrm{occ}}U^{\mathrm{occ}^\dagger})_{ii}\underline{G}_0(\mathrm{i}\tau)_{ij}(U^{\mathrm{occ}}U^{\mathrm{occ}^\dagger})_{jj}\overline{G}_0(\mathrm{i}\tau)_{ab}B^Q_{jb}\\ =& B^P_{\underline{i}\mu}\underline{G}_0(\mathrm{i}\tau)_{\underline{ij}}\overline{G}_0(\mathrm{i}\tau)_{\mu\nu}B^Q_{\underline{j}\nu},\end{aligned} \tag{2.137}$$

and transforming the positive imaginary time Green's function into the AO basis as in the other AO-RPA methods [16, 17]. This technique has reduced the storage requirements by a factor of $N_{\mathrm{bas}}/N_{\mathrm{occ}}$, which allows to treat significantly larger systems. These savings are faciliated by transforming the three center integrals into the Cholesky orbital basis by means of Eq. 2.134 as shown in Eq. 2.137. Therefore, only the more compact AO-CDD three-center integrals have to be stored instead of the entire three-center integrals as in Ref. [17]. An analogous treatment could be carried out for the virtual one-particle density, but as its rank N_{unocc} is typically similar to the basis set dimension, this was omitted for the virtual Green's function in **Publications II, III, V**, because it was found to slightly deteriorate the sparsity pattern [17].
Eq. 2.134 gives a new perspective of CD in correlated electronic structure theory as an orbital rotation similar to the strategies used in local MO techniques. Based on the techniques reviewed in this section and the previous section one can formulate a general strategy to obtain low-scaling electronic structure methods. First, one tries to find a formulation in terms of the non-interacting response function in imaginary time. Next, the arising quantities are transformed to the compact and local CDD-AO basis. The locality of these quantities can then be exploited by the use of sparse-algebra [112, 113]. The ideas illustrated here for the calculation of RPA correlation energies also form the basis for the development of linear- and low-scaling methods to calculate beyond RPA correlation energies and RPA analytical gradients in **Publications III** and **V**.

2.5.4 Low-Scaling Correlated Analytical Gradients

For an electronic structure theory to be broadly applicable to a wide range of problems, properties beyond ground state energies are required. Here, the first step are analytical gradients, which allow, e.g., for geometry optimizations. Computational efficiency is essential for the calculation of molecular properties as well. As the expressions for analytical gradients for an electronic structure method are typically more involved than for the ground state energy, it would be convenient to have a strategy to obtain the working equations in a rather straightforward manner. Here, the ideas that lead to the development of low-scaling analytical RPA gradients in **Publication III** are briefly recapitulated.
Working in the AO-RI framework, the RPA energy can be calculated from the quantities $\mathbf{B}$ (Eq. 2.110), $\mathbf{C}$ (Eq. 2.109), $\mathbf{G}$ (Eq. 2.122) where the non-interacting Green's

function can be calculated purely in the AO basis [100, 114] as

$$\underline{\mathbf{G}}_0(\mathrm{i}\tau) = \mathbf{P}e^{-\tau\mathbf{H}_{\mathrm{KS}}\mathbf{P}} \tag{2.138}$$

$$\overline{\mathbf{G}}_0(\mathrm{i}\tau) = -\mathbf{Q}e^{-\tau\mathbf{H}_{\mathrm{KS}}\mathbf{Q}}. \tag{2.139}$$

Here, the virtual one-particle density matrix is given as

$$Q_{\mu\nu} = \sum_{a}^{\mathrm{unocc}} C_{\mu a}C_{\nu a}. \tag{2.140}$$

As it is well known how to calculate the derivatives of primary and auxiliary basis functions and of the occupied density matrix [101], all contributions can be derived using the chain rule

$$\frac{\partial E_{\mathrm{C}}^{\mathrm{RPA}}}{\partial x} = \frac{\partial E_{\mathrm{C}}^{\mathrm{RPA}}}{\partial \mathbf{C}}\frac{\partial \mathbf{C}}{\partial x} + \frac{\partial E_{\mathrm{C}}^{\mathrm{RPA}}}{\partial \chi_0}\frac{\partial \chi_0}{\partial \mathbf{B}}\frac{\partial \mathbf{B}}{\partial x} + \frac{\partial E_{\mathrm{C}}^{\mathrm{RPA}}}{\partial \mathbf{G}_0}\frac{\partial \mathbf{G}_0}{\partial \mathbf{P}}\frac{\partial \mathbf{P}}{\partial \mathbf{x}}. \tag{2.141}$$

Two of the partial derivatives in Eq. 2.141 correspond to the auxiliary basis representation of important quantitites in MBPT [115]. The partial derivative with respect to the non-interacting response function is known as the correlated Coulomb interaction

$$\mathbf{W}_{\mathrm{C}} = \frac{\partial E_{\mathrm{C}}^{\mathrm{RPA}}}{\partial \chi_0}, \tag{2.142}$$

and the partial derivative with respect to the non-interacting Green's function is known as the self-energy

$$\mathbf{\Sigma}_{\mathrm{C}} = \frac{\partial E_{\mathrm{C}}^{\mathrm{RPA}}}{\partial \mathbf{G}_0}. \tag{2.143}$$

For the calculation of the quantities in Eq. 2.141 the same strategies as described in the previous two sections are applied: performing the rate-determining steps in the imaginary time domain, employing the sparsity of the non-interacting Green's function, and using Cholesky orbitals where possible. This allows to develop asymptotically quadratic scaling RPA analytical gradients as compared to the quartic scaling RI-RPA gradients for molecular systems that were previously the state-of-the-art implementation [18]. The reduced computational requirements of the newly devloped method allows to calculate analytical gradients for systems with more than 600 atoms, significantly extending the reach of first order properties at the RPA level of theory.

To extend this framework to other electron correlation methods one could employ the machinery of MBPT (as detailed in Ref. [62]) to derive the respective expresssions for the correlated Coulomb interaction (Eq. 2.142) and the self-energy (Eq. 2.143). While the derivative w.r.t. the Coulomb operator is not a commonly found quantity in MBPT, it should be possible to use the same techniques to derive this expression.

Chapter 3

Publications

3.1 Publication I: Screening methods for linear-scaling short-range hybrid calculations on CPU and GPU architectures

M. Beuerle, J. Kussmann, C. Ochsenfeld,
"Screening Methods for Linear-Scaling Short-Range Hybrid Calculations on CPU and GPU Architectures",
J. Chem. Phys., **146**, 144108 (2017)

Abstract: We present screening schemes that allow for efficient, linear-scaling short-range exchange calculations employing Gaussian basis sets for both CPU and GPU architectures. They are based on the LinK [C. Ochsenfeld et al., *J. Chem. Phys.* **109**, 1663 (1998)] and PreLinK [J. Kussmann and C. Ochsenfeld, *J. Chem. Phys.* **138**, 134114 (2013)] methods, but account for the decay introduced by the attenuated Coulomb operator in short-range hybrid density functionals. Furthermore, we discuss the implementation of short-range electron repulsion integrals on GPUs. The introduction of our screening methods allows for speedups of up to a factor 7.8 as compared to the underlying linear-scaling algorithm, while retaining full numerical control over the accuracy. With the increasing number of short-range hybrid functionals, our new schemes will allow for significant computational savings on CPU and GPU architectures.

THE JOURNAL OF CHEMICAL PHYSICS **146**, 144108 (2017)

Screening methods for linear-scaling short-range hybrid calculations on CPU and GPU architectures

Matthias Beuerle, Jörg Kussmann, and Christian Ochsenfeld[a)]
Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstrasse 7, D-81377 München, Germany and Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstrasse 5-13, D-81377 München, Germany

(Received 25 December 2016; accepted 28 February 2017; published online 12 April 2017)

We present screening schemes that allow for efficient, linear-scaling short-range exchange calculations employing Gaussian basis sets for both CPU and GPU architectures. They are based on the LinK [C. Ochsenfeld *et al.*, J. Chem. Phys. **109**, 1663 (1998)] and PreLinK [J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013)] methods, but account for the decay introduced by the attenuated Coulomb operator in short-range hybrid density functionals. Furthermore, we discuss the implementation of short-range electron repulsion integrals on GPUs. The introduction of our screening methods allows for speedups of up to a factor 7.8 as compared to the underlying linear-scaling algorithm, while retaining full numerical control over the accuracy. With the increasing number of short-range hybrid functionals, our new schemes will allow for significant computational savings on CPU and GPU architectures. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4978476]

I. INTRODUCTION

Nowadays the most frequently employed method in quantum chemistry is density functional theory (DFT) due to its good compromise between cost and accuracy. Here the choice of an appropriate functional is crucial. A common classification for density functionals is the Jacob's Ladder introduced by Perdew,[1] which categorizes density functionals based on their level of sophistication. In this scheme, level 4 functionals, i.e., hybrid functionals, employ portions of exact exchange in combination with an exchange-correlation functional for the calculation of the exchange correlation energy.[2] Among these are the highly cited B3LYP[2–5] and PBE0[6] functionals as examples for the high success of this class of functionals. A variant of the PBE0 functional is the HSE06 functional,[7–9] which employs short-range exact exchange introduced through the substitution of the Coulomb operator $1/r_{12}$ with the attenuated Coulomb operator $\mathrm{erfc}(\omega r_{12})/r_{12}$. The introduction of the attenuated Coulomb operator reduces computational cost and enables a more elaborate combination of density functional based exchange and exact exchange, by describing the short-range exchange with exact exchange and resorting to a density functional description in the long range regime. This is especially appealing not only for solid state applications but also for calculations of large molecules. In solid state applications the attenuated Coulomb operator decreases the number of unit cells that have to be included in the exact exchange calculation to obtain converged results.[10] Other functionals employing screened exchange have also been introduced.[11,12] A very promising short-range density functional with dispersion correction, HSE-3c, was recently introduced by Grimme *et al.*,[13] with the intention to provide a good description of non-covalent interactions for molecular crystals and molecules at low computational cost. We note that there are also long-range corrected DFT functionals, i.e., functionals that combine short-range DFT exchange with long range exact exchange.[14,15] The purpose of those functionals is to correct the wrong long-range exchange behavior of short-range and "conventional" hybrid density functionals. A more elaborate scheme to combine density functional and exact exchange at different ranges is the Coulomb-attenuating method.[16]

While the accuracy of DFT calculations is generally increased by employing exact exchange, the computational cost also increases since the formation of the exchange matrix in the basis of the Kohn-Sham orbitals becomes the rate determining step. To reduce the cost of the exchange matrix formation, algorithms have been introduced to achieve a linear-scaling formation on conventional CPU[17–20] and more recently on GPU architectures.[21] We employ the LinK method,[17,18] which has been highly successful for the exchange formation on CPUs using a sophisticated loop structure in combination with one particle density matrix weighted Schwarz integral estimates[22] to achieve linear scaling. A straightforward extension of the LinK procedure on GPUs is not possible since it requires branching, which leads to different execution paths within a warp. This is prohibitive on a GPU architecture. Therefore, the PreLinK[21] method was developed, which also employs the one particle density matrix and Schwarz integral estimates to preselect significant exchange matrix elements prior to the calculation of the two electron integrals. While using the LinK and the PreLinK methods for screened hybrid DFT calculations is possible, the additional decay introduced by the attenuated Coulomb operator is not explicitly taken into account, which leaves room for further improvement. A distance dependent screening routine employing integral estimates for the attenuated Coulomb operator in combination

a)christian.ochsenfeld@uni-muenchen.de

with a fast multipole method (FMM)-like boxing scheme has been presented for screened hybrid calculations.[23] Since the LinK and PreLinK routines have proven to be highly efficient, linear-scaling routines customized to the formation of the exact exchange matrix, adaption to short-range hybrids should enable fast, linear-scaling short-range hybrid DFT calculations with decreased prefactor. In this work, we describe modifications of the LinK and PreLinK methods in combination with novel screening criteria inspired by QQR estimates,[24] taking into account the additional decay, introduced by the attenuated Coulomb operator. We show that this allows for efficient short-range hybrid calculations for large molecules on CPU and GPU computing architectures employing gaussian basis sets.

II. THEORY

A. ERFC LinK screening

Density functionals relying on exact exchange calculations require the formation of the exchange matrix in the basis of Kohn-Sham orbitals.[2] In a conventional exchange matrix formation, the elements $K_{\mu\nu}$ are calculated in an atomic orbital basis as follows:

$$K_{\mu\nu} = \sum_{\lambda,\sigma} P_{\lambda\sigma}\,(\mu\lambda|\nu\sigma)\,, \tag{1}$$

where the $(\mu\lambda|\nu\sigma)$ terms correspond to electronic repulsion integrals (ERIs) of the Coulomb operator $1/r_{12}$ over the distributions $\Omega_{\mu\lambda}$ and $\Omega_{\nu\sigma}$. Contraction with the respective element of the one-particle density matrix $P_{\lambda\sigma}$ yields the exchange matrix element $K_{\mu\nu}$. A naive calculation of $\mathbf{K}$ would therefore scale as $\mathcal{O}(\mathrm{N}^4)$, where N is the number of basis functions. This can be reduced to $\mathcal{O}(\mathrm{N}^2)$ when employing Schwarz integral estimates,[22]

$$|\,(\mu\lambda|\nu\sigma)\,| \leq Q_{\mu\lambda}Q_{\nu\sigma} = \sqrt{(\mu\lambda|\mu\lambda)}\sqrt{(\nu\sigma|\nu\sigma)}. \tag{2}$$

In the case of electronic insulators, the coupling of the bra- and ket-distributions in Eq. (1) with the local one-particle density matrix $P_{\lambda\sigma}$ leads to a further reduction of significant integrals that have to be evaluated. This is exploited in a number of algorithms to obtain an overall linear-scaling formation of the exchange matrix $\mathbf{K}$ (e.g., Ref. 17). In the LinK scheme,[17,18] significant ERIs are selected using density weighted Schwarz integral estimates, which ensures a linear-scaling behavior of the algorithm for insulators, while preserving the highly optimized structure of integral evaluation schemes.

In electronic structure calculations using short-range hybrids,[7] the Coulomb operator $1/r_{12}$ is substituted by the attenuated Coulomb operator $\mathrm{erfc}(\omega r_{12})/r_{12}$, when forming the short-range exchange matrix $\mathbf{K}'$. This leads to the following integrals:

$$(\mu\lambda|\nu\sigma) = \iint \frac{\chi_\mu(r_1)\chi_\lambda(r_1)\mathrm{erfc}(\omega r_{12})\chi_\nu(r_2)\chi_\sigma(r_2)}{r_{12}} dr_1 dr_2, \tag{3}$$

which have to be evaluated during the short-range exchange matrix formation. Here ω is a screening parameter, which depends on the functional used in combination with the screened exchange contribution. A common value is $\omega = 0.11$ used in, e.g., the HSE06, HSE-3c, and N12-SX functionals.[9,11,13]

When performing a screened hybrid calculation with the conventional LinK method, the additional decay introduced by the attenuated Coulomb operator can be employed during the integral evaluation by comparing the intermediate T_ω to a given threshold (see Sec. III), to evaluate at an early stage, whether the integral is significant or not.[25] As will be shown in the present work, there is, however, significant capacity to modify the LinK method to further reduce the computational effort, i.e., to yield a linear-scaling formation of the exchange matrix with reduced prefactor. This is achieved by modifying the screening criteria in the original LinK method, to account for the additional decay introduced by the attenuated Coulomb operator $\mathrm{erfc}(\omega r_{12})/r_{12}$, which leads to the sketch of the ERFC LinK method shown in Figure S1 of the supplementary material. The adapted method resembles the original LinK loop structure but employs additional screening criteria to reduce the screening effort (pre-sorting loop) and the number of integrals that have to be evaluated (screening loop). In the pre-sorting routine, $R_{\mathrm{min},\mu\nu}$ is defined as

$$R_{\mathrm{min},\mu\nu} = R_{\mu\nu} - ext_\mu - ext_\nu \tag{4}$$

using the reasoning that the extent of shell μ is an upper bound to the extent of all significant shellpairs $\Omega_{\mu\lambda}$. Because the complementary error function is a strictly decaying function, the following inequality holds:

$$\left|\int\int \frac{\chi_\mu(r_1)\chi_\lambda(r_1)\mathrm{erfc}(\omega r_{12})\chi_\nu(r_2)\chi_\sigma(r_2)}{r_{12}}\,\mathrm{d}r_1\mathrm{d}r_2\right| \geq \mathrm{erfc}(\omega R_{\mathrm{min},\mu\nu})|(\mu\lambda|\nu\sigma)| \geq \mathrm{erfc}(\omega R_{\mathrm{min},\mu\nu})Q_{\mu\lambda}Q_{\nu\sigma}. \tag{5}$$

Furthermore, we introduce QQR integral estimates $Q_{\mu\lambda}Q_{\nu\sigma}/R_{\mathrm{min},\mu\nu}$,[24] to take into account the inherent decay of the $1/r_{12}$ operator. The relationship between this approach and combining the decay of the attenuated Coulomb operator[23] with Schwarz integral estimates[22] is discussed below.

The extents ext_μ are calculated as the maximum extent of all primitive basis functions in each shell μ, where

the definition from Sierka *et al.*[26] was employed for ext_μ,

$$ext_\mu = \max_{\forall\, \alpha \in \mu} \left\{ \sqrt{\frac{-\ln \varepsilon + \frac{1}{2} \ln \zeta_{\mu\alpha}}{\zeta_{\mu\alpha}}} \right\}, \tag{6}$$

where $\zeta_{\mu\alpha}$ corresponds to the exponent of the primitive basis function α and ε corresponds to a pre-defined threshold. We note that in Eqs. (6) and (9) the primitive contraction coefficients are not taken into account. While it would be theoretically possible to incorporate contraction coefficients by, e.g., weighting the threshold or multiplying the extents with a logarithmic expression containing the contraction coefficients, we decided to use established shell extents, which have proven to be useful in earlier studies.[24,26,27] Introducing this modified pre-screening routine reduces the number of terms for which the computationally demanding erfc-function has to be evaluated in the screening routine. In the pre-screening routine, introducing a modified screening criterion is, however, not essential. Numerical tests have shown that simply using the conventional Schwarz estimate in the pre-screening loop also yields good results since the major speedup is obtained in the screening loop.

In the screening routine (shellpair loop), we employ the following criteria using the same ideas that lead to Eq. (5):

$$|P_{\lambda\sigma}(\mu\lambda|\nu\sigma)| \approx |P_{\lambda\sigma}| \cdot \sqrt{(\mu\lambda|\mu\lambda)} \cdot \sqrt{(\nu\sigma|\nu\sigma)} \cdot \frac{\operatorname{erfc}(\omega R_{\mu\lambda,\nu\sigma})}{R_{\mu\lambda,\nu\sigma}}, \tag{7}$$

where $R_{\mu\lambda,\nu\sigma}$ is defined as

$$R_{\mu\lambda,\nu\sigma} = R_{PQ} - ext_{\mu\lambda} - ext_{\nu\sigma}, \tag{8}$$

where R_{PQ} corresponds to the distance between the centers of the charge distributions $\Omega_{\mu\lambda}$ and $\Omega_{\nu\sigma}$ centered at **P** and **Q**. The shellpair-centers **P** (and **Q**) were defined according to Maurer *et al.*[24] For the extent $ext_{\mu\lambda}$ of the distribution $\Omega_{\mu\lambda}$, the definitions employed for QQR estimates,[24] which are inspired by the well-seperated extent of the continuous fast multipole method (CFMM),[27] are used. The explicit definition is as follows:

$$ext_{\mu\lambda} = \max_{\forall\, \alpha,\beta \,in\, \mu,\lambda} \left\{ \sqrt{\frac{2}{\zeta_{\mu\alpha} + \zeta_{\lambda\beta}}} \operatorname{erfc}^{-1}(\vartheta_{\text{thr}}) + r_{\mu\lambda,\alpha\beta} \right\}, \tag{9}$$

where again $\zeta_{\mu\alpha}$ and $\zeta_{\lambda\beta}$ correspond to the exponents of the respective primitive basis function and ϑ_{thr} corresponds to a predefined threshold. The term $r_{\mu\lambda,\alpha\beta}$ corrects for the deviation of the center of the primitive shellpair to the center of the contracted shellpair as defined by Maurer *et al.*[24] The maximum of all primitive extent is used for each shellpair. Subtracting the shellpair extent is essential to guarantee the validity of Eq. (5) and the QQR estimates.[24]

An alternative screening criterion can be derived using the decay of the attenuated Coulomb operators, which was given in this context by Izmaylov *et al.*[23] as

$$(\mu\lambda|\nu\sigma) \sim \frac{\operatorname{erfc}(\theta_\omega R_{\mu\lambda,\nu\sigma})}{R_{\mu\lambda,\nu\sigma}}, \tag{10}$$

where θ_ω^2 is given by

$$\theta_\omega^2 = \min_{\forall\, \alpha,\beta,\gamma,\delta \in \mu,\nu,\lambda,\sigma} \times \left\{ \frac{1}{(\zeta_{\mu\alpha} + \zeta_{\lambda\beta})^{-1} + (\zeta_{\nu\gamma} + \zeta_{\sigma\delta})^{-1} + \omega^{-2}} \right\}. \tag{11}$$

In our context we combine Eq. (10) with Schwarz integral estimates[22] to obtain

$$|P_{\lambda\sigma}(\mu\lambda|\nu\sigma)| \approx |P_{\lambda\sigma}| \cdot \sqrt{(\mu\lambda|\mu\lambda)} \cdot \sqrt{(\nu\sigma|\nu\sigma)} \cdot \frac{\operatorname{erfc}(\theta_\omega R_{\mu\lambda,\nu\sigma})}{R_{\mu\lambda,\nu\sigma}} \tag{12}$$

with $R_{\mu\lambda,\nu\sigma}$ defined as in Eq. (8). Minimal values of θ_ω are used to ensure validity for all primitive ERIs. We found that both our screening criteria give highly similar results, which suggests that

$$\theta_\omega = \sqrt{\frac{1}{(\zeta_{\mu\alpha} + \zeta_{\lambda\beta})^{-1} + (\zeta_{\nu\gamma} + \zeta_{\sigma\delta})^{-1} + \omega^{-2}}} \approx \omega. \tag{13}$$

This can be understood since for the basis sets tested here (def2-SVP, def2-SVPD, def2-TZVP) the minimal exponents $\zeta_{\mu\alpha}$ are on the same order of magnitude as ω. Therefore, the ω term is dominant since it appears squared compared to the exponents. Both criteria have been extensively tested and yield similar accuracy with slightly increased efficiency of the criterion shown in Eq. (7).

B. ERFC PreLinK pre-screening

While the LinK method has been highly successful on CPU architectures, it is not compatible with massively parallel computing architectures such as GPUs. The reason for this is that the branching inherent to the LinK method strongly reduces the efficiency on GPUs. Therefore, the PreLinK[21] scheme has been developed to ensure a linear-scaling formation of the exchange matrix on GPUs. In the Pre-LinK scheme[21] a screening matrix $\mathbf{Q}'$ is calculated prior to the actual calculation of the exchange matrix **K**, to allow for a linear-scaling formation of the exchange matrix suitable for use on GPUs. The elements $Q'_{\mu\nu}$ represent an upper bound to $K_{\mu\nu}$ and are calculated as follows:

$$Q'_{\mu\nu} = \sum_{\lambda,\sigma} |P_{\lambda\sigma}| \sqrt{(\mu\lambda|\mu\lambda)} \sqrt{(\nu\sigma|\nu\sigma)} \geq K_{\mu\nu}. \tag{14}$$

The entire matrix $\mathbf{Q}'$ is obtained by two matrix multiplications,

$$\mathbf{Q}' = \mathbf{Q} \times |\mathbf{P}| \times \mathbf{Q}, \tag{15}$$

which can be performed in a linear-scaling manner using sparse algebra routines. The essential characteristic of this screening method is that the significant matrix elements are selected prior to the calculation of the two electron integrals, which allows efficient computation of the exchange matrix on massively parallel computing architectures.

The PreLinK screening is the basis for a modified procedure applicable to more precisely pre-select exchange matrix

elements in short-range screened hybrid DFT calculations suitable for massively parallel computing architectures. Examining the screening criterion of the ERFC LinK routine (Eq. (7)) and setting $R_{\mu\lambda,\nu\sigma}$ constant to a value smaller or equal to its minimal value $R_{\min,\mu\nu}$ for all significant distributions $\Omega_{\mu\lambda}$, $\Omega_{\nu\sigma}$ appearing in the calculation of $K_{\mu\nu}$ allows to reformulate Eq. (7) using Eqs. (14) and (15) as

$$K'_{\mu\nu} \approx \frac{\mathrm{erfc}(\omega R_{\min,\mu\nu})}{R_{\min,\mu\nu}} Q'_{\mu\nu}. \tag{16}$$

This can be used for screening purposes by equating the RHS of Eq. (16) to a threshold $\vartheta_{\mathrm{erfc}}$,

$$\vartheta_{\mathrm{erfc}} = \frac{\mathrm{erfc}(\omega R_{\min,\mu\nu})}{R_{\min,\mu\nu}} Q'_{\mu\nu}. \tag{17}$$

If the RHS of Eq. (17) is smaller than the threshold, one does not have to calculate the element $\mathbf{K}'$. For the threshold $\vartheta_{\mathrm{erfc}}$, the most natural choice is to simply employ the thresholds recommended for the conventional PreLinK-scheme.[21] $R_{\min,\mu\nu}$ is defined as in the pre-screening routine of the ERFC Link routine (see Eq. (4)).

Alternatively one can also employ the decay of the attenuated Coulomb operator[23] (in analogy to Eq. (12)) and combine this with the PreLinK scheme to yield

$$K'_{\mu\nu} \approx \frac{\mathrm{erfc}(\theta_{\omega,\min} R_{\min,\mu\nu})}{R_{\min,\mu\nu}} Q'_{\mu\nu}. \tag{18}$$

A minimal value for $\theta^2_{\omega,\min}$ is given by

$$\theta^2_{\omega,\min} = \frac{1}{(\zeta_{\mu\alpha,\min} + \zeta_{\lambda\beta,\min})^{-1} + (\zeta_{\nu\gamma,\min} + \zeta_{\sigma\delta,\min})^{-1} + \omega^{-2}}, \tag{19}$$

where $\zeta_{\mu\alpha,\min}$ and $\zeta_{\nu\gamma,\min}$ correspond to the minimal exponents of shell μ and ν, and $\zeta_{\lambda\beta,\min}$ and $\zeta_{\sigma\delta,\min}$ are set to the minimal exponents of the basis set used for the given molecule. As for the ERFC LinK routine, we have tested both criteria and again both have shown to give similar accuracy. The criterion in Eq. (16) has shown to yield increased speedups, which is why solely these results are shown in Sec. IV.

In the implementation of the ERFC PreLinK method, we start with conventional PreLinK pre-screening since setting up $\mathbf{Q}'$ is required for Eq. (16) and comparing each matrix element $Q'_{\mu\nu}$ against a pre-screening threshold ϑ_{pre} is cheap and reduces the number of terms for which Eq. (17) has to be evaluated. The remaining screening procedure described above then reduces the prefactor of the linear-scaling formation of $\mathbf{K}'$ for screened hybrid calculations, facilitated by the introduction of the attenuated Coulomb operator $\mathrm{erfc}(\omega r_{12})/r_{12}$.

C. T_ω-Pre-Screening

Since the results obtained with the ERFC PreLinK screening are not as satisfactory as for the CPU based ERFC LinK routine, we also investigated a different approach to test whether the performance of the GPU based screening algorithms can be improved. For this approach, we examined the fundamental integrals that occur during the evaluation of short-range ERIs (Eq. (3)). We mainly follow the notation in Ref. 28. The fundamental integrals are evaluated using the following intermediates:[25]

$$\sigma_P = \frac{1}{\zeta_{\mu\alpha} + \zeta_{\lambda\beta}}, \quad \sigma_Q = \frac{1}{\zeta_{\nu\gamma} + \zeta_{\sigma\delta}}, \tag{20}$$

$$\mathbf{P} = (\zeta_{\mu\alpha}\mathbf{A} + \zeta_{\lambda\beta}\mathbf{B})\sigma_P, \quad \mathbf{Q} = (\zeta_{\nu\gamma}\mathbf{C} + \zeta_{\sigma\delta}\mathbf{D})\sigma_Q, \tag{21}$$

$$\begin{aligned} U_P &= (\pi\sigma_P)^{3/2} D_A D_B e^{-\zeta_{\mu\alpha}\zeta_{\lambda\beta}\sigma_P |\mathbf{A}-\mathbf{B}|^2}, \\ U_Q &= (\pi\sigma_Q)^{3/2} D_C D_D e^{-\zeta_{\nu\gamma}\zeta_{\sigma\delta}\sigma_Q |\mathbf{C}-\mathbf{D}|^2}, \end{aligned} \tag{22}$$

$$R^2 = |\mathbf{P} - \mathbf{Q}|^2, \tag{23}$$

$$\theta^2 = \frac{1}{\sigma_P + \sigma_Q}, \quad \theta^2_\omega = \frac{1}{\sigma_P + \sigma_Q + \frac{1}{\omega^2}}, \tag{24}$$

$$T = \theta^2 R^2, \quad T_\omega = \theta^2_\omega R^2, \tag{25}$$

which are combined as follows to give the $[\mathbf{0}]^{(m)}$:

$$[\mathbf{0}]^{(m)} = U_P U_Q \{(2\theta^2)^{m+1/2} G_m(T) - (2\theta^2_\omega)^{m+1/2} G_m(T_\omega)\}, \tag{26}$$

where $\zeta_{\mu\alpha}$, $\zeta_{\lambda\beta}$, $\zeta_{\nu\gamma}$, $\zeta_{\sigma\delta}$ are the exponents corresponding to the primitive basis functions μ, λ, ν, σ located at centers $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ with coefficients D_A, D_B, D_C, D_D. $G_m(T)$ represents the Boys function[29,30] of argument T scaled by $(2/\pi)^{1/2}$. Now it is well established that if T_ω exceeds some critical value T_{crit}, the integral $(\mu\lambda|\nu\sigma)$ is negligible and does not need to be calculated.[25] Therefore if the minimum value of T_ω of all $(\mu\lambda|\nu\sigma)$ appearing in the calculation of $K'_{\mu\nu}$ is known and when this value $T_{\omega,\min}$ exceeds T_{crit}, this particular element of the exchange matrix $K'_{\mu\nu}$ does not have to be evaluated. Minimum values for θ^2_ω for shells μ and ν can be obtained as described in Sec. II B. T_{crit} is calculated as

$$T_{\mathrm{crit}} = -2 \log \vartheta_{\mathrm{T}}, \tag{27}$$

where ϑ_{T} corresponds to an accuracy threshold usually set to $\vartheta_{\mathrm{int}}/100$.

In the actual screening routine, one then performs a PreLinK pre-screening first, similar to the routine described in Sec. II B. For the remaining elements $K_{\mu\nu}$, one evaluates $\theta^2_{\omega,\min}$ and solves for $R_{\mathrm{cut},\mu\nu}$, i.e., the biggest separation of two shells μ and ν for them to still lead to a significant interaction $(\Omega_{\mu\lambda}|\Omega_{\nu\sigma})$. The explicit definition of $R_{\mathrm{cut},\mu\nu}$ is given by

$$R_{\mathrm{cut},\mu\nu} = \sqrt{\frac{T_{\mathrm{crit}}}{\theta^2_{\omega,\min}}} + ext_\mu + ext_\nu. \tag{28}$$

To ensure that $R_{\mathrm{cut},\mu\nu}$ is valid for all significant charge distributions including shells μ and ν, we add the extents of shells μ and ν to $R_{\mathrm{cut},\mu\nu}$. If $R_{\mu\nu}$ for the element $\mathrm{K}_{\mu\nu}$ is bigger than $R_{\mathrm{cut},\mu\nu}$, one does not have to explicitly evaluate $\mathrm{K}_{\mu\nu}$. Since we found that this procedure leads to a systematic overestimation of $R_{\mathrm{cut},\mu\nu}$, an empirical scaling factor is applied to $R_{\mathrm{cut},\mu\nu}$ to obtain a compromise between accuracy and speed. The empirical scaling parameter x will be denoted as $\mathrm{T}^{\mathrm{x}}_\omega$-screening or skipped if the default parameter is used (see also Sec. III).

As mentioned above, pre-selection of significant integrals is required, which is why rather crude estimates have

to be used on massively parallel computing architectures. For the PreLinK scheme, this is due to the fact that absolute values for the one particle density matrix **P** have to be used.[21] For the additional screening routines accounting for the attenuated Coulomb operator described above, this is due to the need to use conservative screening quantities to retain accuracy.

III. COMPUTATIONAL DETAILS

A. CPU implementation

The evaluation of short-range ERIs on CPUs was performed using the Obara-Saika scheme.[31] The only difference compared to the evaluation of conventional ERIs is the need to calculate the scaled Boys Function G_m twice (see Eq. (26)), namely, once for the argument T and once for T_ω.[25] This also enables the introduction of an early exit criterion $T_\omega \geq T_{\text{crit}}$ (see Eq. (25)), which reduces the computational workload without any additional screening.[25] The recurrence relations employed in the Obara-Saika scheme do not change. Since especially for large l-quantum number combinations the recurrence relations are computationally dominant, the evaluation of short-range ERIs on CPUs proceeds almost as fast as for conventional ERIs.

B. GPU implementation

The implementation of the formation of the short-range exact exchange matrix is based on the implementation for the conventional exchange matrix formation routine on GPUs.[21,32] All GPU calculations were performed using a double precision implementation. We employ both the McMurchie Davidson (McD) scheme[33] and Rys quadrature[34,35] for the evaluation of the electronic repulsion integrals (ERIs) over the attenuated Coulomb operator, depending on the l-quantum number combination of the basis functions. In the McD scheme the main difference between the evaluation of the ERIs over the attenuated Coulomb operator and the conventional ERIs is that the calculation of the fundamental integrals $[0]^m$ requires approximately double the effort since two Boys Kernel for T and T_ω have to be evaluated for each $m \leq L$ (see Eq. (26)). The evaluation of the ERIs then proceeds through the application of recurrence relations to the $[0]^m$ to obtain the final result. Since especially for large l-quantum numbers the recurrence relations take up most of the computational time, the time required for the evaluation of the ERI over the attenuated Coulomb operator is approximately the same as for conventional ERIs when the McD scheme is used. As will be described in the next paragraph, this is a good argument for the use of the McD scheme compared to Rys quadrature. On GPUs the use of the McD scheme, however, becomes prohibitive when dealing with large l-quantum numbers due to the high memory per core requirements of this integral evaluation scheme.[21]

The integral evaluation using Rys quadrature employs a numeric quadrature to evaluate the two terms in Eq. (26). More precisely, a quadrature of the polynomial t^{2m}, with t as the integration variable of the Boys function of order m and the remainder of the expression constituting the weighting function, is performed using Rys polynomials. Since the weighting function depends on T or T_ω, respectively, one has to perform one complete quadrature for each term, which results in approximately double the computational cost for the attenuated ERIs as compared to conventional ERIs. This limits the computational benefit of short-range hybrid functionals on GPUs, when integral evaluation is performed using Rys quadrature, since even though less integrals have to be evaluated, the evaluation of short-range ERIs using Rys quadrature is significantly more expensive than the evaluation of conventional ERIs. Despite this limitation Rys quadrature becomes more efficient than the McD scheme for the evaluation of short-range ERIs with large l-quantum numbers, due to the high memory requirements of the McD scheme in those cases. Another performance limiting factor is the reduced efficiency of the early exit condition described in Sec. III A on GPUs due to the requirement to synchronize all threads involved in the calculation of a primitive exchange matrix element.

As described in the previous paragraphs, choosing the integral evaluation scheme for short-range ERIs on GPUs merely based on Flop count considerations is not possible since the performance is mainly memory bound. For the same reasons, Kussmann and Ochsenfeld[21] timed each l-quantum number combination to find the optimal balance between Rys quadrature and the McD scheme for computational efficiency. Due to the differences in the evaluation of short-range and conventional ERIs using Rys quadrature, those considerations are not optimal for the present purpose. Therefore we again timed the performance of the McD scheme vs. Rys quadrature for each l-quantum combination performing one short-range exchange matrix formation for a DNA_{16} molecule and an entire SCF-calculation for $Amylose_8$ and DNA_4 using the def2-SVP basis set. Based on these calculations (NVIDIA GeForce GTX Titan GPU device), we find the McD scheme to be more efficient for the following l-quantum combinations:

$$[ss, ss], [sp, sp], [sp, ps], [sd, ss],$$

$$[sd, sp], [sd, sd], [sd, ps], [sd, pp],$$

$$[sd, ds], [ps, ss], [ps, ps], [pp, ss],$$

$$[pp, sp], [pp, ps], [pp, pp], [pd, sp],$$

$$[pd, ps], [ds, ss], [ds, sp], [ds, ps],$$

$$[ds, pp], [ds, ds], [dp, sp], [dp, ps].$$

For the remaining l-quantum number combinations and all l-quantum combinations containing $l \geq 3$, we use Rys quadrature. The ERFC PreLinK and T_ω-Pre-Screening routines were implemented prior to the integral evaluation on CPUs since the computational requirements for the screening are negligible.

C. Calculations and timings

We implemented all screening routines and performed all calculations in the FermiONs++ program package developed in our group.[21,36] The integral kernels were compiled using

the NVIDIA 7.5 toolkit. All CPU timings were recorded on 12 core CPU servers (2x Intel Xeon CPU E5645 @ 2.40 GHz). In the case of GPU calculations, one NVIDIA GTX Titan was used in combination with a 12 core CPU server (2x Intel Xeon CPU E5-2620 v2 @ 2.10 GHz).

As an initial guess, we used a superposition of atomic densities, generated for each individual atom by an atomic SCF calculation prior to the actual SCF calculation. As a convergence accelerator we used Pulay's DIIS[37,38] and Saunders-Hillier level shifting.[39] The exchange-correlation energy provided by the relevant functional was calculated using the LIBXC library.[40] We employed a 99 (radial) 590 (angular) Lebedev/Laikov-grid, with an M4 radial grid[41] for all DFT calculations. The integral screening threshold ϑ_{int} was set to 10^{-10}. For the determination of the extents, ϵ (Eq. (6)) was set to 10^{-9}, ϑ_{thr} (Eq. (9)) was set to 10^{-1}, and ϑ_{T} was set to 10^{-12}. A convergence criterion (FPS-commutator) of $\vartheta_{\text{conv}} = 10^{-7}$ was used in all calculations unless stated otherwise. We employed the continuous fast multipole method (CFMM)[27] to allow for a linear scaling calculation of the Coulomb energy on CPUs. On GPUs the Coulomb matrix formation was performed based on the McD algorithm[33] in analogy to the J-engine method[42,43] adapted for massively parallel computing architectures.[32] The empirical weighting parameter used to scale R_{crit} in the T_ω-Pre-Screening routine was set to 0.20 based on test calculations on a small set of molecules unless stated otherwise.

Information about molecular geometries and all total energies and timings are presented in the supplementary material. The reported speedups correspond to the fraction of time used for the conventional LinK/PreLinK formation of the exchange matrix accumulated during the SCF cycle divided by the time used for the exchange matrix formation including the relevant screening procedure. It has to be noted that the early exit criterion $T_\omega \geq T_{\text{crit}}$ is employed in both the reference and screened calculations. This entails that the reported speedups are caused only by the introduction of the relevant screening routine. The errors were calculated for the final SCF energy compared to the unscreened calculation employing a threshold of $\vartheta_{\text{CPU}} = 10^{-10}$ used in the LinK and ERFC LinK routines and $\vartheta_{\text{GPU}} = 10^{-4}$ used in the PreLinK and ERFC PreLinK/T_ω-Pre-Screening calculations.

IV. RESULTS AND DISCUSSION

A. General purpose test using the HSE06 functional

As a first test, we evaluated the performance of our screening routines on a large portion of the test-set introduced for the benchmarking of the QQR integral estimates.[24] The test-set covers a variety of large molecules with different chemical properties to ensure robustness along a wide range of molecular systems. We tested the ERFC LinK routine using both the criteria from Eqs. (7) and (12) to show the high similarity between the two. Furthermore we evaluated the performance of the GPU based ERFC PreLinK (criterion from Eq. (16)) and T_ω-Pre-Screening routine.

The results of this benchmark (Table I) show that the errors introduced by the ERFC LinK routines are negligible (below 10^{-7} H). Furthermore Table I shows that, as stated in Sec. II, both criteria (Eqs. (7) and (12)) employed in the screening routine show similar results in terms of speedups and accuracy. Therefore solely the results obtained with the criterion from Eq. (7) will be shown in the subsequent tests. The highest speedup observed for the CPU routine was 7.8 for Polyene_{1024}. It has to be noted that some of the systems in Table I are not single reference systems and would have to be treated with multireference techniques, to obtain physically reliable results. Here they serve, however, as test cases for electronically delocalized systems.[24] Other notable examples are the water clusters and the carbon nanotubes, which show considerable speedups of up to a factor 3.8 for H_2O_{569} and 5.0 for $\text{CNT}(6, 3)_8$. In the specific case of $\text{CNT}(6, 3)_8$ (690 atoms, def2-SVP basis,[44] ~9500 basis functions), introducing the ERFC LinK routine leads to a reduction of the total wall time required for exchange calculations from 179 h to 36 h, while the total SCF wall time reduces from 205 h to 61 h. While traditionally the formation of the exchange matrix is the most time consuming step, employing the ERFC LinK routine renders the computational cost of this step more similar, or even smaller than the formation of the Coulomb matrix in most cases. This is illustrated in a scatter plot (see Figure 1), which shows the computational cost required for the exact exchange part vs. the Coulomb part for the test set in Table I. The average ratio of the time required for the formation of the exchange matrix compared to the Coulomb matrix $T_{tot,K}/T_{tot,J}$ is 3.4 when employing the conventional LinK algorithm and 1.3 when employing the ERFC LinK scheme. This shows that DFT calculations employing short-range exact exchange can be performed with significantly reduced computational cost on CPU architectures.

For the GPU based ERFC PreLinK screening, the errors also do not exceed 10^{-7} H. The maximum speedup observed for ERFC PreLinK screening is 2.07. The average speedup for the entire test set with the ERFC PreLinK routine is ~1.4 alongside with a negligible average error (~7 *n*hartree). For the T_ω-Pre-Screening routine, the errors are significantly higher compared to both the ERFC PreLinK and the ERFC LinK routine. The maximum error is 56 μhartree (0.04 kcal/mol) for polyene_{1024} and the average error is ~4 μhartree. It has to be noted that the higher average error as compared to the ERFC LinK and ERFC PreLinK method is mainly accounted for by systems with small HOMO-LUMO gap. For the remaining systems, the errors are below 1 μhartree. While the errors are higher as compared to the ERFC PreLinK routine, the speedups are also higher with a maximum speedup of 3.2 for the carbon nanotube $\text{CNT}(6, 3)_8$ and an average speedup of ~1.8 across the entire test set. Even though the speedups for the formation of the exchange matrix on GPUs using the ERFC PreLinK and the T_ω-Pre-Screening routine are significantly lower as compared to the ERFC LinK scheme, the impact on the wall time required for the total SCF calculation for the GPU routines is more similar to the ERFC LinK scheme. This is due to the fact that the proportion of time required for exchange calculations in comparison to the total time required for the SCF calculation is higher on GPUs. In the case of DNA_{16} (1052 atoms, def2-SVP basis,[44] ~11 000 basis functions) a speedup in the exchange part of 3.2 translates into a total speedup in wall time of 2.2 for the ERFC LinK scheme on CPUs, while a

TABLE I. Speedups (S.U.) and errors for the different screening routines for DFT calculations using the HSE06 functional[7,9] with the def2-SVP basis set.[44] Speedups correspond to the fraction of time required for exchange matrix formations during the complete SCF cycle employing the reference algorithm (LinK for CPUs and PreLinK for GPUs), divided by the time required employing the additional screening routine. Errors are reported for total, converged SCF energies relative to the reference calculation.

		ERFC LinK @CPU							
		Eq. (12)		Eq. (7)		ERFC PreLinK @GPU		Tω-pre-screening @GPU	
System	N_{Atoms}	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.
$Amylose_2$	45	0.2	1.14	0.2	1.13	0.3	1.02	0.5	1.04
$Amylose_4$	87	0.5	1.42	0.5	1.42	0.4	1.14	1.4	1.33
$Amylose_8$	171	1.2	1.71	1.2	1.72	0.7	1.30	3.9	1.61
$Amylose_{16}$	339	2.2	1.84	2.0	1.84	1.3	1.35	5.8	1.71
$Amylose_{32}$	675	4.0	1.87	3.7	1.84	0.4	1.36	8.7	1.73
$Amylose_{48}$	1011	5.3	1.85	5.7	1.86	1.3	1.36	14.1	1.72
$Amylose_{64}$	1347	5.0	1.85	5.5	1.86	0.4	1.35	20.1	1.69
Beta-carotene	96	0.3	1.87	0.3	1.89	1.8	1.58	1123.2	2.09
CNT^{a}_{20}	30	0.1	1.04	0.2	1.04	0.0	0.96	0.0	0.98
CNT^{a}_{40}	50	1.4	1.11	1.5	1.12	0.2	1.0	299.7	1.01
CNT^{a}_{80}	90	0.1	1.37	0.2	1.38	1.1	1.03	3116.7	1.27
CNT^{a}_{160}	170	0.1	2.10	1.0	2.12	5.6	1.38	13277.8	2.08
$CNT(6,3)^{a}_{8}$	690	58.9	4.91	65.6	4.97	78.5	1.89	40221.7	3.22
$Diamond_{102}$	102	1.2	1.15	1.3	1.14	1.2	1.00	0.6	1.00
DNA_1	62	0.1	1.44	0.1	1.40	0.1	1.19	5.2	1.40
DNA_2	128	0.1	1.62	0.0	1.64	0.2	1.23	14.2	1.49
DNA_4	260	0.5	2.09	0.5	2.11	0.0	1.39	35.1	1.90
DNA_8	524	2.1	2.74	2.2	2.78	1.9	1.62	88.1	2.42
DNA_{16}	1026	4.0	3.20	4.0	3.23	2.7	1.71	161.4	2.61
$Graphite_{24}$	36	0.3	1.11	0.3	1.12	0.7	1.00	13.4	1.03
$Graphite_{54}$	72	3.6	1.30	3.5	1.31	0.7	1.01	4337.0	1.20
$Graphite_{96}$	120	9.7	1.57	9.0	1.59	5.5	1.08	8487.5	1.58
$(H_2O)_{68}$	204	0.3	1.94	0.3	1.96	3.1	1.33	2.3	1.50
$(H_2O)_{142}$	426	1.5	2.54	1.4	2.55	9.0	1.51	1.2	1.84
$(H_2O)_{285}$	855	2.7	3.17	3.5	3.19	19.2	1.67	6.8	2.17
$(H_2O)_{569}$	1707	7.8	3.76	9.3	3.80	53.5	1.78	35.7	2.41
$(LiF)_{32}$	32	0.5	1.17	0.7	1.14	0.0	1.03	127.5	1.14
$(LiF)_{72}$	72	5.0	1.26	5.5	1.28	0.3	1.03	432.3	1.19
$Polyethyne_{64}$	130	0.4	2.54	0.3	2.58	10.7	1.88	719.6	2.55
$Polyethyne_{128}$	258	0.8	3.29	0.7	3.36	23.2	1.96	1509.2	2.65
$Polyene_{64}$	66	2.0	3.15	2.5	3.11	0.1	1.85	3290.2	3.10
$Polyene_{1024}$	1026	36.2	7.77	40.6	7.81	0.5	2.07	56122.2	2.40
$(S_8)_5$	40	1.3	1.51	1.2	1.55	0.9	1.24	138.3	1.44
$(S_8)_{20}$	160	6.3	2.57	6.4	2.60	10.2	1.71	972.7	2.38
Average		4.9	2.21	5.3	2.22	6.9	1.38	3958.6	1.79

[a]Due to convergence problems a SCF energy difference criterion ϑconv = 10^{-6} was used.

a speedup in the exchange part of 2.6 leads to a total speedup of 2.0 for the T$_\omega$-Pre-Screening routine on GPUs.

To test the accuracy of our routines on a large, covalent, globular system, we performed a calculation on the crystal structure of a 56 amino acid long protein domain (DGCR8 Dimerization Domain, PDB-ID: 4E5R,[45] 975 atoms, def2-SVP basis,[44] 9176 basis functions). The ERFC LinK routine yields a speedup of a factor of 3.0 for the exchange matrix calculations, while the error remains below 1 *n*hartree. The GPU based ERFC PreLink method leads to a speed up of 1.7, while the error remains well below 10 *n*hartree. The T$_\omega$-Pre-Screening routine leads to a speedup of 2.8 along with an error of 91 *n*hartree. These results show that also for large, covalent, globular systems our screening routines remain accurate, while yielding significant speedups. This is a further confirmation that weighting Schwarz integral estimates with the decay of the operator at hand in a QQR-type fashion,[24] while not mathematically rigorous, is a viable approach to accurate and efficient screening methods.

While the numbers from Table I and the more tedious integral evaluation for ERIs over the attenuated Coulomb operator make the GPU routines appear less attractive, it is enlightening to compare the wall times of a calculation employing 12 CPU cores to the time required for the same calculation on a single GPU. The scatter plot in Figure 2 shows the wall times required for the entire SCF calculations for all systems in Table I on a 12 core CPU server compared to the wall times when both the Coulomb and exchange matrix calculations are

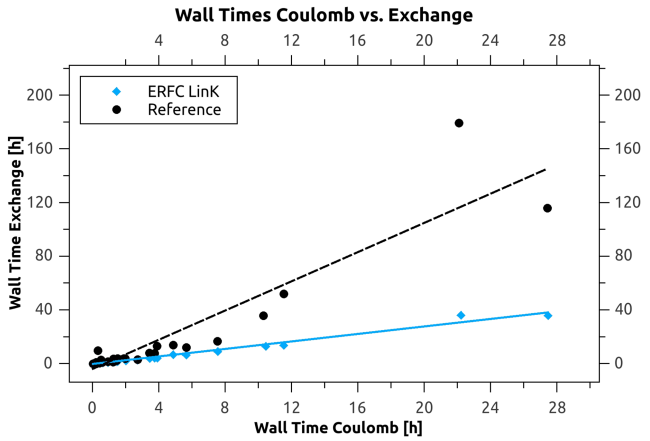


FIG. 1. Scatter plot showing the accumulated wall times required for the formation of the short-range exact exchange matrix compared to the wall time required for the formation of the Coulomb matrix for each system in Table I during the entire SCF cycle on CPUs. Shown are the numbers for the LinK reference and the ERFC LinK routine employing the screening criterion from Eq. (7).

performed on one GPU. On average the speedup obtained with the T_ω-Pre-Screening routine compared to the ERFC LinK calculations was 5.9, while the ERFC PreLinK calculations were 4.7x faster. Therefore when both implementations are available and the use of a short-range hybrid is required due to its special features or as a reference, the GPU code is recommended.

B. Scaling behavior

To make sure that our screening routines do not change the linear-scaling behavior of the underlying linear-scaling algorithms for the formation of the exchange matrix, we performed calculations on a set of linear alkanes ($Alkane_X$, X = 1-320). Such almost one dimensional systems are ideal to show the asymptotic scaling behavior of a quantum chemical method since the asymptotic limit is reached with modest computational requirements.[46] Furthermore these calculations serve as a test, whether the objective, namely, linear-scaling routines with reduced prefactor on CPUs and GPUs, have been obtained. The results for both the CPU and GPU routines are shown in Figure 3. It is evident that all routines achieve the aim of providing a reduced prefactor linear-scaling formation of the short-range exchange matrix. Furthermore the benefit of the additional screenings remains constant with increasing system size in the asymptotic limit. The highest prefactor reduction is achieved by the ERFC LinK routine followed by the T_ω-Pre-Screening and ERFC PreLinK routines, which is in line with the average speedups seen in Table I.

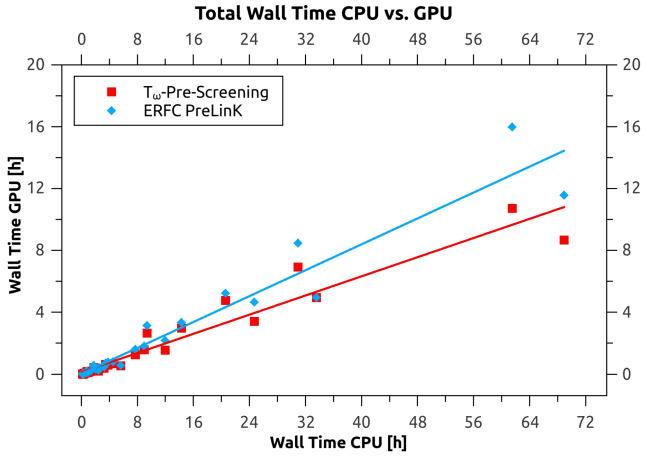


FIG. 2. The graph shows the total wall times required for an entire SCF cycle on a 12 core CPU server compared to the use of one GPU for the Coulomb and exchange integral routines for each system in Table I. Both the T_ω-Pre-Screening and the ERFC PreLinK GPU routine are compared to the ERFC LinK timings. To indicate a rough trend, linear fits to each data set are shown as well.

C. Basis set dependence

Furthermore we examined the basis set dependence of our screening routines from small (def2-SV) to large (def2-TZVP) and also the influence of diffuse basis functions (def2-SVPD).[44] For calculations with large and diffuse basis sets, we focused on a subset of the molecules used in Table I. The results are shown in Table II.

For the ERFC LinK routine, the speedups increase in the following order def2-SV < def2-SVP < def2-TZVP, while the errors stay on the same order of magnitude. The speedups

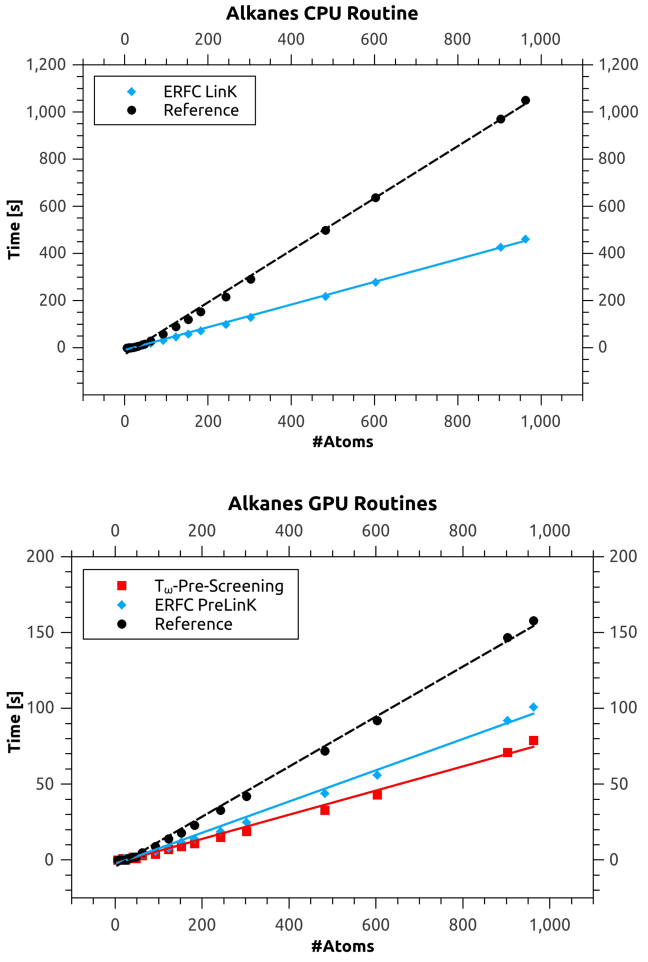


FIG. 3. The graphs show the time required for the formation of one short-range exact exchange matrix with a converged one particle density matrix plotted against the number of atoms in the system. The results for the CPU routines are shown in the top and the results for the GPU routines are shown in the bottom graph.

TABLE II. Speedups (S.U.) and errors for the ERFC LinK, ERFC PreLinK and T_ω-Pre-Screening routines for DFT calculations using the HSE06 functional[7,9] with different def2-X (X = SV, SVP, SVPD, TZVP) basis sets.[44] Speedups correspond to the fraction of time required for exchange matrix formations during the complete SCF cycle employing the reference algorithm (LinK for CPUs and PreLinK for GPUs), divided by the time required employing the additional screening routine. Errors are reported for total, converged SCF energies relative to the reference calculation.

ERFC LinK@CPU									
		def2-SV		def2-SVP		def2-TZVP		def2-SVPD	
System	N_{Atoms}	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.
$Amylose_8$	171	1.4	1.65	1.2	1.72	0.2	2.04	3.5	1.85
DNA_2	128	0.1	1.54	0.0	1.64	0.1	1.74	1.9	1.47
$(H_2O)_{68}$	204	0.1	1.84	0.3	1.96	0.6	2.21	1.0	1.69
$Polyethyne_{128}$	258	0.9	2.77	0.7	3.36	2.3	4.30	6.8	3.86
$(S_8)_5$	40	2.5	1.54	1.2	1.55	0.6	1.64	2.7	1.32
Average		1.0	1.87	0.68	2.05	0.8	2.39	3.2	2.04
ERFC PreLinK@GPU									
		def2-SV		def2-SVP		def2-TZVP		def2-SVPD	
System	N_{Atoms}	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.
$Amylose_8$	171	0.3	1.28	0.7	1.30	0.3	1.46	3.0	1.19
DNA_2	128	0.0	1.20	0.2	1.23	0.8	1.15	1.5	1.03
$(H_2O)_{68}$	204	3.1	1.26	3.1	1.33	4.3	1.39	1.8	1.02
$Polyethyne_{128}$	258	22.5	1.84	23.2	1.96	10.9	2.13	0.6	1.61
$(S_8)_5$	40	0.7	1.26	0.9	1.24	3.2	1.12	0.9	1.03
Average		5.3	1.37	5.62	1.41	3.9	1.45	1.6	1.18
T_ω-pre-screening@GPU									
		def2-SV		def2-SVP		def2-TZVP		def2-SVPD	
System	N_{Atoms}	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.
$Amylose_8$	171	3.2	1.61	3.9	1.61	22.6	2.14	53.2	2.02
DNA_2	128	10.9	1.48	14.2	1.49	20.1	1.53	11.1	1.43
$(H_2O)_{68}$	204	1.8	1.41	2.3	1.50	5.4	1.68	5.6	1.39
$Polyethyne_{128}$	258	338.6	2.54	1509.2	2.65	11479.3	3.28	621.8	2.82
$(S_8)_5$	40	135.6	1.48	138.3	1.44	93.9	1.44	126.1	1.25
Average		98.0	1.70	333.6	1.74	2324.5	2.01	163.6	1.78

obtained when including diffuse functions (def2-SVPD) do not show a clear trend; however, the error remains on the same order of magnitude. A similar behavior is observed for the ERFC PreLinK scheme; however, the errors for the calculations employing the def2-SVPD basis set are decreased. The same holds true for the T_ω-Pre-Screening routine, with errors more or less independent of the basis set used. This shows that all three routines deliver accurate results independent of the basis set. The highest speedups were obtained with the def2-TZVP basis, enabling fast short-range exchange calculations employing a triple zeta basis. This indicates that the speedups seen in Table I, e.g., 7.8, for polyene 1024, are expected to increase, when using a bigger basis set.

D. Different ω-screening parameters

Finally we examined the performance of our screening routines using a short-range density functional with a different ω-value than HSE06. Here, Moussa *et al.*[12] have shown that $\omega = 0.22$ (with different mixing parameter $a = 0.425$) yields similar results in terms of accuracy as HSE06. This is appealing since a higher ω-value entails reduced computational cost due to increased screening of the exchange contribution. The results for the ERFC LinK and the GPU screening routines are shown in Table III.

For the ERFC LinK scheme, the speedups obtained with $\omega = 0.22$ are higher than the ones obtained for $\omega = 0.11$ in all cases with a maximum increase of a factor of 1.8 in the case of $(S_8)_{20}$. At the same time, the errors obtained with the ERFC LinK routine decrease. The average increase in speedup for the molecules shown in Table III for $\omega = 0.22$ compared to $\omega = 0.11$ was 1.4. The total wall time for the exchange matrix calculations decreases from 65 min to 38 min for $(S_8)_{20}$ (same number of SCF iterations), highlighting the potential of even higher speedups, when using short-range hybrid functionals with higher ω-parameters.

For the ERFC PreLinK, no significantly increased speedups were obtained on average for the higher ω-value. Similar to the ERFC LinK routine, the errors decrease when changing ω to 0.22. The calculation with T_ω-Pre-Screening shows both significant speedups and increase in errors, which indicates that the scaling factor applied to R_{cut} to yield a certain accuracy is ω dependent. Performing the calculations with a scaling factor of 0.25 instead of 0.20 reduces the error

TABLE III. Speedups (S.U.) and errors for the ERFC LinK, ERFC PreLinK, and T_ω-Pre-Screening routines for DFT calculations using the HSE12s functional[12] ($\omega = 0.22$) with the def2-SVP basis set.[44] Speedups correspond to the fraction of time required for exchange matrix formations during the complete SCF cycle employing the reference algorithm (LinK for CPUs and PreLinK for GPUs), divided by the time required employing the additional screening routine. Errors are reported for total, converged SCF energies relative to the reference calculation.

		ERFC LinK @CPU		ERFC PreLinK @GPU		$T_\omega^{0.20}$-Pre-Screening @GPU		$T_\omega^{0.25}$-Pre-Screening @GPU	
System	N_{Atoms}	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.	E[10^{-9}H]	S.U.
$Amylose_8$	171	0.5	2.24	0.2	1.37	2188.9	2.65	3.0	2.05
DNA_2	128	0.1	2.15	0.5	1.29	1056.8	2.62	14.7	1.91
$(H_2O)_{68}$	204	0.1	2.90	0.2	1.43	1859.3	3.46	7.9	2.31
$Polyethyne_{128}$	258	0.2	4.45	0.0	1.61	39460.4	2.72	477.0	2.26
$Polyene_{1024}$	1026	5.6	9.50	0.1	1.94	n.c.[a]	n.c.[a]	14383.0	2.14
$(S_8)_5$	40	1.1	2.26	0.0	1.34	242.3	2.69	27.5	2.00
$(S_8)_{20}$	160	6.0	4.59	0.3	1.87	10228.7	6.05	179.2	3.84
Average $\omega = 0.22$		1.9	4.01	0.2	1.55	9172.7	3.37	2156.0	2.36
Average $\omega = 0.11$		7.2	2.95	5.5	1.55	8394.7	1.92	571.8	1.55

[a]n.c.: Calculation did not converge because of too extensive screening.

significantly, but obviously also the speedups obtained. While this ω dependence of the scaling factor is not optimal, calibration on a small test set for the ω value at hand should yield a reliable value quickly. Furthermore, most short-range functionals employ $\omega = 0.11$, for which the values employed in the majority of this study show stable behavior of the T_ω-Pre-Screening routine.

To test the ω-dependence of the screening routines across a wider ω-range, we performed calculations on $(H_2O)_{68}$ keeping the exchange mixing parameter fixed at the HSE value of 0.25, but changing ω from 0.03 to 0.9. The results in Figure 4 show that for the ERFC LinK routine the speedup increases steadily with increasing ω up to a maximum of approximately 7. For the T_ω-Pre-Screening routines with different scaling parameters, the overall trend is the same for each scaling parameter $T_\omega^{0.20-0.30}$, namely, an increase in speedup up to $\omega = 0.4$ and then approximately constant behavior. Obviously a lower weighting parameter shows a higher speedup. The ERFC PreLinK routine shows a slight increase in speedup up to $\omega = 0.2$ and then constant behavior.

When examining the ω-dependence of the errors introduced by the screening, Figure 4 shows that the error for the ERFC LinK routine does not significantly exceed 10^{-9} H across the entire ω-range. For the ERFC PreLinK routine, the error decreases with increasing ω and never exceeds 10^{-7} H showing that both the ERFC LinK and ERFC PreLinK routine do not cause significant errors across the entire ω-range. For the T_ω-Pre-Screening graphs the error increases similar to the speedup with increasing ω until $\omega = 0.4$, after which the error approximately remains constant (two points at $\omega = 0.3$ and $\omega = 0.6$ deviate from the overall trend on the curve for the $T_\omega^{0.20}$- and $T_\omega^{0.25}$-Pre-Screening, respectively). As mentioned earlier this behavior is not satisfactory; however, the error can be easily reduced, when resorting to a higher scaling parameter at higher ω-values. For the scaling parameter 0.20, the error raises above the acceptable error of approximately 1 μhartree at $\omega = 0.25$. The error when employing $T_\omega^{0.25}$-Pre-Screening does however not significantly exceed 1 μhartree, and the errors for $T_\omega^{0.30}$-Pre-Screening are even lower.

A study by Moussa *et al.*,[12] examined the HSE parameter space by varying the exact exchange mixing parameter and the screening parameter ω. The authors[12] found that

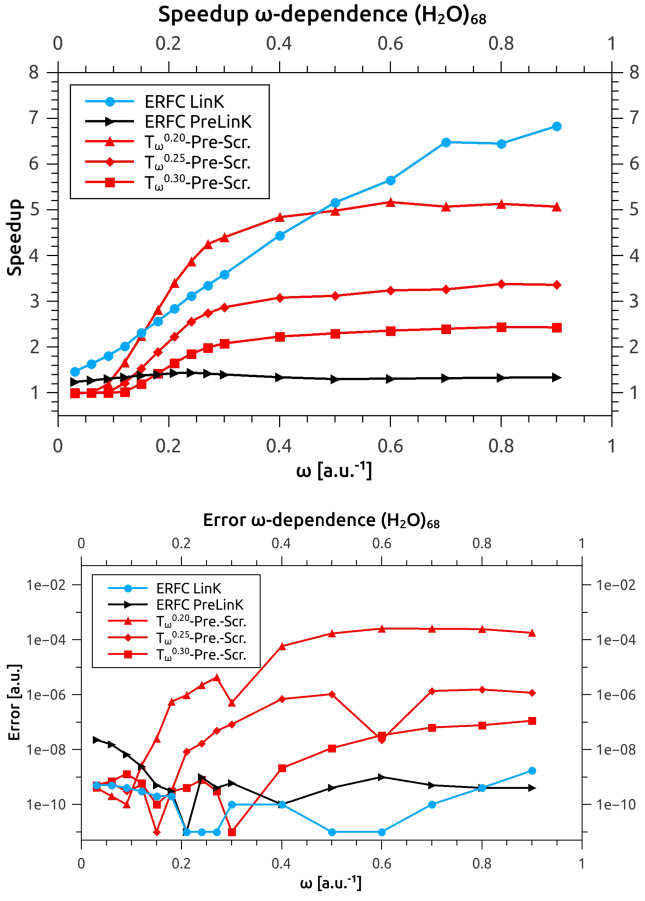


FIG. 4. The top graph shows a plot of the speedup of the wall time required for the exchange matrix formation during the entire SCF calculation as compared to the reference implementation (LinK for ERFC LinK, PreLink for T_ω-Pre-Screening and ERFC PreLinK) against the screening parameter ω employed in the calculation. The bottom graph shows a plot of the total SCF error as compared to the reference implementation against the screening parameter ω employed in the calculation. The superscript of the T_ω-Pre-Screening labels indicates the empirical scaling parameter used for the calculation of the matrix element estimates.

for $\omega > 0.22$ no functional can be built, which yields similar accuracy as compared to HSE06. This means that the study of the behavior of our screening routines for $\omega > 0.22$ is mainly conceptual until another density functional exchange description is available, which enables accurate results for higher ω-values. For the relevant ω-space, all of our screening routines show satisfactory results in terms of accuracy and speedup.

V. CONCLUSION

We have presented reduced prefactor linear-scaling algorithms for short-range exchange matrix calculations. The CPU based ERFC LinK routine yields speedups of up to a factor of 7.8, while the error introduced by the screening is not significant. Among the GPU screening routines, the ERFC PreLinK shows lower speedups. The performance-increase obtained with the T_ω-Pre-Screening method comes, however, with an increased error as compared to the ERFC PreLinK method. Higher errors were observed especially for higher ω-values and low HOMO-LUMO gap systems. We showed, however, that the error can be easily controlled by an empirical scaling parameter. The speedups obtained with the GPU based routines are in general lower as compared to the ERFC LinK routine since a more crude preselection of matrix elements is necessary on massively parallel computing architectures.[21]

We also discussed the problems involved with the implementation of short-range ERIs on GPUs. While this makes short-range exchange calculations on GPUs seem less attractive, we showed that compared to the CPU implementation, significant speedups are nevertheless possible.

While the proof of concept calculations shown in this article focused on molecular systems, short-range hybrid calculations are especially popular in the solid state community. We are convinced that an extension of our methodology developed here is readily applicable to periodic boundary condition calculations employing Gaussian basis sets, since all necessary quantities are available.

Compared to an earlier study on efficient short-range exchange calculations,[23] we have introduced a new screening criterion in the case of the T_ω-Pre-Screening routine and combined different ideas from general screening considerations[23,24,47] to introduce screening criteria compatible with the LinK and PreLinK schemes. Furthermore, our linear-scaling algorithms do not depend on a boxing scheme, which enables a more fine-grained screening and is also readily implemented in pre-existing LinK and PreLinK implementations. While for periodic boundary conditions, boxes (unit cells) are naturally present, we believe that the ERFC LinK, ERFC PreLinK, and T_ω-Pre-Screening routine can also be beneficial here. This is especially true for systems, which require large unit cells.

SUPPLEMENTARY MATERIAL

See supplementary material for molecular geometries, total energies, and timings.

ACKNOWLEDGMENTS

C.O. acknowledges financial support from the Excellence Cluster EXC114 "Center for Integrated Protein Science Munich" (CIPSM) and SFB749.

[1] J. P. Perdew and K. Schmidt, AIP Conf. Proc. **577**, 1 (2001).
[2] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[3] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[4] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[5] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[6] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).
[7] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).
[8] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **124**, 219906 (2006).
[9] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, J. Chem. Phys. **125**, 224106 (2006).
[10] J. Heyd and G. E. Scuseria, J. Chem. Phys. **121**, 1187 (2004).
[11] R. Peverati and D. G. Truhlar, Phys. Chem. Chem. Phys. **14**, 16187 (2012).
[12] J. E. Moussa, P. A. Schultz, and J. R. Chelikowsky, J. Chem. Phys. **136**, 204117 (2012).
[13] J. G. Brandenburg, E. Caldeweyher, and S. Grimme, Phys. Chem. Chem. Phys. **18**, 15519 (2016).
[14] T. Leininger, H. Stoll, H.-J. Werner, and A. Savin, Chem. Phys. Lett. **275**, 151 (1997).
[15] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).
[16] T. Yanai, D. P. Tew, and N. C. Handy, Chem. Phys. Lett. **393**, 51 (2004).
[17] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[18] C. Ochsenfeld, Chem. Phys. Lett. **327**, 216 (2000).
[19] E. Schwegler and M. Challacombe, J. Chem. Phys. **105**, 2726 (1996).
[20] E. Schwegler, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[21] J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013).
[22] M. Haeser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[23] A. F. Izmaylov, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **125**, 104103 (2006).
[24] S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, J. Chem. Phys. **136**, 144107 (2012).
[25] R. D. Adamson, J. P. Dombroski, and P. M. W. Gill, J. Comput. Chem. **20**, 921 (1999).
[26] M. Sierka, A. Hogekamp, and R. Ahlrichs, J. Chem. Phys. **118**, 9136 (2003).
[27] C. A. White, B. G. Johnson, P. M. Gill, and M. Head-Gordon, Chem. Phys. Lett. **230**, 8 (1994).
[28] P. M. W. Gill, B. G. Johnson, and J. A. Pople, Int. J. Quantum Chem. **40**, 745 (1991).
[29] S. F. Boys, Proc. R. Soc. A **200**, 542 (1950).
[30] P. M. W. Gill, Adv. Quantum. Chem. **25**, 141 (1994).
[31] S. Obara and A. Saika, J. Chem. Phys. **84**, 3963 (1986).
[32] I. S. Ufimtsev and T. J. Martinez, J. Chem. Theory Comput. **5**, 1004 (2009).
[33] L. E. McMurchie and E. R. Davidson, J. Comput. Phys. **26**, 218 (1978).
[34] H. F. King and M. Dupuis, J. Comput. Phys. **21**, 144 (1976).
[35] J. Rys, M. Dupuis, and H. F. King, J. Comput. Chem. **4**, 154 (1983).
[36] J. Kussmann and C. Ochsenfeld, J. Chem. Theory Comput. **11**, 918 (2015).
[37] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[38] P. Pulay, J. Comput. Chem. **3**, 556 (1982).
[39] V. R. Saunders and I. H. Hillier, Int. J. Quantum Chem. **7**, 699 (1973).
[40] M. A. Marques, M. J. Oliveira, and T. Burnus, Comput. Phys. Commun. **183**, 2272 (2012).
[41] O. Treutler and R. Ahlrichs, J. Chem. Phys. **102**, 346 (1995).
[42] C. A. White and M. Head-Gordon, J. Chem. Phys. **104**, 2620 (1996).
[43] G. R. Ahmadi and J. Almloef, Chem. Phys. Lett. **246**, 364 (1995).
[44] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[45] R. Senturia, A. Laganowsky, I. Barr, B. D. Scheidemantle, and F. Guo, PLoS One **7**, e39688 (2012).
[46] H. F. Schurkus and C. Ochsenfeld, J. Chem. Phys. **144**, 031101 (2016).
[47] S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **138**, 014101 (2013).

Supplementary Material for Screening Methods for Linear-Scaling Short-Range Hybrid Calculations on CPU and GPU Architectures

Matthias Beuerle, Jörg Kussmann, and Christian Ochsenfeld[a)]

Chair of Theoretical Chemistry,

Department of Chemistry,

University of Munich (LMU),

Butenandtstr. 7, D-81377 München, Germany

[a)] christian.ochsenfeld@uni-muenchen.de

Contents

1 ERFC LinK Scheme

loop over types (angular momenta, contraction,...) of shell-pair blocks
  loop over all μ's in significant bra-shell pairs
    loop over all ν's in significant ket-shell pairs
      if $([|\mathbf{P}_{\mu\nu}| \cdot (\mu_{\max}|\mu_{\max})^{1/2} \cdot (\nu_{\max}|\nu_{\max})^{1/2}] >$ Threshold) **then**
        if $([|\mathbf{P}_{\mu\nu}| \cdot (\mu_{\max}|\mu_{\max})^{1/2} \cdot (\nu_{\max}|\nu_{\max})^{1/2} \cdot \frac{\mathrm{erfc}(\omega \mathrm{R}_{\min,\mu\nu})}{\mathrm{R}_{\min,\mu\nu}}] >$ Threshold) **then**
          Store significant ν's for each μ
        end if
      end if
    end loop
    Sort ν's by size of $[|\mathbf{P}_{\mu\nu}| \cdot (\nu_{\max}|\nu_{\max})^{1/2}]$ for each μ
  end loop
  loop over significant bra-shell pairs $\mu\lambda$
    loop over significant ν's corresponding to μ
      loop over significant σ's
        if $([|\mathbf{P}_{\mu\nu}| \cdot (\mu\lambda|\mu\lambda)^{1/2} \cdot (\nu\sigma|\nu\sigma)^{1/2}] >$ Threshold) **then**
          $\mathrm{ctr}_\sigma + +$
          if $([|\mathbf{P}_{\mu\nu}| \cdot (\mu\lambda|\mu\lambda)^{1/2} \cdot (\nu\sigma|\nu\sigma)^{1/2} \cdot \frac{\mathrm{erfc}(\omega \mathrm{R}_{\mu\lambda,\nu\sigma})}{\mathrm{R}_{\mu\lambda,\nu\sigma}}] >$ Threshold) **then**
            Add $(\nu\sigma)$ to ML_μ
          end if
        else Leave σ-loop
        end if
      end loop
      if ctr_σ is zero **then** leave σ-loop
    end loop
    Same loop for λ to form ML_λ
    Merging sorted shell-pair lists ML_μ and ML_λ yields ML
    loop over significant ket-shell pairs $\nu\sigma \in$ ML
      Form $(\mu\lambda|\nu\sigma)$ and contract with $\mathbf{P}_{\mu\nu}$, $\mathbf{P}_{\nu\sigma}$, $\mathbf{P}_{\lambda\nu}$, $\mathbf{P}_{\lambda\sigma}$
    end loop
  end loop
end loop

Figure S1: Pseudocode of the ERFC LinK, scheme resembling the original LinK loop structure [1], with additional screening criteria to account for the decay introduced by the attenuated Coulomb operator.

2 Molecular Structures

The structures for the test set used in this article were adopted from Maurer et al. [2] and are available on our group home-page along with the linear alkanes [3].

The crystal-structure of the DGCR8 dimerization domain was downloaded from the PDB (PDB-ID: 4E5R [4]).

Subsequently missing hyrdogens were added using Avogadro [5].

3 Total Energies and Timings

3.1 CPU calculations

3.1.1 Data for Section 'General Purpose Test using the HSE06 functional'

Table 1: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on CPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
Amylose_{2}	-1295.4567121864	1187	525
Amylose_{4}	-2514.6590442323	3703	1943
Amylose_{8}	-4953.0635615979	9541	5630
Amylose_{16}	-9829.8723965973	22186	13815
Amylose_{32}	-19583.4903964247	45086	28128
Amylose_{48}	-29337.1094519594	71645	43765
Amylose_{64}	-39090.7291669768	101972	60696
Beta-carotene	-1555.0156201718	3013	1721
CNT^{*}_{20}	-766.3696901960	578	294
CNT^{*}_{40}	-1526.9882093163	3479	1846
CNT^{*}_{80}	-3048.2929933866	13775	8133
CNT^{*}_{160}	-6090.8435904133	69452	50361
$\text{CNT}(6,3)^{*}_{8}$	-25576.2260942159	736787	645817
Diamond_{102}	-1631.7173281050	21524	10988
DNA_{1}	-1760.3240502280	1639	864
DNA_{2}	-4502.3006369713	11520	6655
DNA_{4}	-9986.2377491667	42942	29091
DNA_{8}	-20954.1262332797	170854	128900
DNA_{16}	-42889.9083699813	535800	417343
Graphite_{24}	-920.2459654458	915	468
Graphite_{54}	-2064.5666929172	5687	3504
Graphite_{96}	-3665.9943514948	18509	13196
$(\text{H}_2\text{O})_{68}$	-5188.2204447163	6438	3755
$(\text{H}_2\text{O})_{142}$	-10834.5826300924	20628	13703
$(\text{H}_2\text{O})_{285}$	-21746.0223160767	66161	47589
$(\text{H}_2\text{O})_{569}$	-43416.3054753086	249328	187652
$(\text{LiF})_{32}$	-1717.1001635839	548	263
$(\text{LiF})_{72}$	-3863.8868816667	9673	4883
Polyethyne_{64}	-2473.5765328628	2047	1335
Polyethyne_{128}	-4945.9875084588	5244	3607
Polyyne_{64}	-2433.2728417950	771	510
Polyyne_{1024}	-38914.7735893304	151322	35327
$(\text{S}_8)_5$	-15918.0655865763	689	440
$(\text{S}_8)_{20}$	-63672.3245839813	12482	10122
DGCR8 Dimerization Domain	-24313.7175998831	310205	234829

Table 2: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC Link (Eq. 12) calculations on CPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
Amylose_{2}	-1295.4567121862	1122	461
Amylose_{4}	-2514.6590442318	3133	1372
Amylose_{8}	-4953.0635615967	7233	3295
Amylose_{16}	-9829.8723965951	15914	7521
Amylose_{32}	-19583.4903964207	32089	15071
Amylose_{48}	-29337.1094519541	51333	23657
Amylose_{64}	-39090.7291669718	74119	32731
Beta-carotene	-1555.0156201715	2214	921
CNT^{*}_{20}	-766.3696901961	572	282
CNT^{*}_{40}	-1526.9882093149	3302	1669
CNT^{*}_{80}	-3048.2929933867	11538	5948
CNT^{*}_{160}	-6090.8435904132	43146	23990
$\text{CNT}(6,3)^{*}_{8}$	-25576.2260942748	222623	131633
Diamond_{102}	-1631.7173281062	20087	9593
DNA_{1}	-1760.3240502281	1371	602
DNA_{2}	-4502.3006369714	9026	4096
DNA_{4}	-9986.2377491662	27699	13909
DNA_{8}	-20954.1262332776	88992	47060
DNA_{16}	-42889.9083699773	249036	130469
Graphite_{24}	-920.2459654461	868	423
Graphite_{54}	-2064.5666929136	4884	2699
Graphite_{96}	-3665.9943514851	13733	8425
$(\text{H}_2\text{O})_{68}$	-5188.2204447160	4638	1940
$(\text{H}_2\text{O})_{142}$	-10834.5826300909	12347	5389
$(\text{H}_2\text{O})_{285}$	-21746.0223160740	33520	15033
$(\text{H}_2\text{O})_{569}$	-43416.3054753008	111490	49949
$(\text{LiF})_{32}$	-1717.1001635834	511	224
$(\text{LiF})_{72}$	-3863.8868816617	8655	3869
Polyethyne_{64}	-2473.5765328624	1233	525
Polyethyne_{128}	-4945.9875084580	2727	1097
Polyyne_{64}	-2433.2728417930	420	162
Polyyne_{1024}	-38914.7735892942	120760	4547
$(\text{S}_8)_5$	-15918.0655865776	535	291
$(\text{S}_8)_{20}$	-63672.3245839876	6296	3938

Table 3: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC LinK (Eq. 7) calculations on CPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
Amylose_{2}	-1295.4567121862	1118	463
Amylose_{4}	-2514.6590442318	3145	1372
Amylose_{8}	-4953.0635615967	7185	3266
Amylose_{16}	-9829.8723965953	16055	7502
Amylose_{32}	-19583.4903964210	32167	15261
Amylose_{48}	-29337.1094519537	51292	23498
Amylose_{64}	-39090.7291669713	73875	32562
Beta-carotene	-1555.0156201715	2200	911
CNT^{*}_{20}	-766.3696901962	568	282
CNT^{*}_{40}	-1526.9882093148	3285	1655
CNT^{*}_{80}	-3048.2929933864	11506	5908
CNT^{*}_{160}	-6090.8435904143	42910	23781
$\text{CNT}(6,3)^{*}_{8}$	-25576.2260942815	221204	129867
Diamond_{102}	-1631.7173281063	20104	9601
DNA_{1}	-1760.3240502281	1387	618
DNA_{2}	-4502.3006369713	8918	4067
DNA_{4}	-9986.2377491662	27667	13778
DNA_{8}	-20954.1262332775	88839	46400
DNA_{16}	-42889.9083699773	247986	129260
Graphite_{24}	-920.2459654461	869	419
Graphite_{54}	-2064.5666929137	4862	2685
Graphite_{96}	-3665.9943514858	13612	8296
$(\text{H}_2\text{O})_{68}$	-5188.2204447160	4614	1918
$(\text{H}_2\text{O})_{142}$	-10834.5826300910	12299	5366
$(\text{H}_2\text{O})_{285}$	-21746.0223160732	33568	14939
$(\text{H}_2\text{O})_{569}$	-43416.3054752993	111167	49431
$(\text{LiF})_{32}$	-1717.1001635832	513	230
$(\text{LiF})_{72}$	-3863.8868816612	8648	3822
Polyethyne_{64}	-2473.5765328625	1228	517
Polyethyne_{128}	-4945.9875084581	2702	1074
Polyyne_{64}	-2433.2728417925	423	164
Polyyne_{1024}	-38914.7735892898	120669	4525
$(\text{S}_8)_5$	-15918.0655865775	529	284
$(\text{S}_8)_{20}$	-63672.3245839877	6251	3893
DGCR8 Dimerization Domain	-24313.7175998833	153300	77678

3.1.2 Data for Section 'Scaling Behavior'

Table 4: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on CPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Alkane_{01}$	-40.4316067044	2	≤ 1
$Alkane_{02}$	-79.6689867297	6	≤ 1
$Alkane_{03}$	-118.9088500395	14	4
$Alkane_{04}$	-158.1482944884	25	8
$Alkane_{05}$	-197.3876326675	46	16
$Alkane_{06}$	-236.6269683354	69	26
$Alkane_{07}$	-275.8663020060	104	43
$Alkane_{08}$	-315.1056240125	176	54
$Alkane_{09}$	-354.3449491534	229	77
$Alkane_{10}$	-393.5842687765	270	93
$Alkane_{13}$	-511.3022317827	407	166
$Alkane_{15}$	-589.7808721998	496	222
$Alkane_{20}$	-785.9774723309	717	360
$Alkane_{30}$	-1178.3706721881	1240	691
$Alkane_{40}$	-1570.7638713537	1835	1081
$Alkane_{50}$	-1963.1570704472	2422	1458
$Alkane_{60}$	-2355.5502699057	3030	1838
$Alkane_{80}$	-3140.3366684876	4303	2618
$Alkane_{100}$	-3925.1230669812	5676	3496
$Alkane_{160}$	-6279.4822619087	10173	6035
$Alkane_{200}$	-7849.0550585895	13929	7674
$Alkane_{300}$	-11772.9870495731	25945	11815
$Alkane_{320}$	-12557.7734475081	29640	12692

Table 5: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC LinK (Eq. 7) calculations on CPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Alkane_{01}$	-40.4316067044	3	≤ 1
$Alkane_{02}$	-79.6689867297	6	1
$Alkane_{03}$	-118.9088500395	14	3
$Alkane_{04}$	-158.1482944884	25	7
$Alkane_{05}$	-197.3876326674	46	17
$Alkane_{06}$	-236.6269683354	68	26
$Alkane_{07}$	-275.8663020061	102	41
$Alkane_{08}$	-315.1056240125	172	50
$Alkane_{09}$	-354.3449491534	226	70
$Alkane_{10}$	-393.5842687765	261	86
$Alkane_{13}$	-511.3022317828	379	138
$Alkane_{15}$	-589.7808721998	448	175
$Alkane_{20}$	-785.9774723309	598	243
$Alkane_{30}$	-1178.3706721881	936	390
$Alkane_{40}$	-1570.7638713537	1307	555
$Alkane_{50}$	-1963.1570704472	1669	705
$Alkane_{60}$	-2355.5502699057	2052	864
$Alkane_{80}$	-3140.3366684875	2877	1203
$Alkane_{100}$	-3925.1230669811	3729	1552
$Alkane_{160}$	-6279.4822619082	6794	2642
$Alkane_{200}$	-7849.0550585889	9607	3349
$Alkane_{300}$	-11772.9870495723	19340	5209
$Alkane_{320}$	-12557.7734475072	22544	5622

3.1.3 Data for Section 'Basis Set Dependence'

Table 6: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on CPUs using the HSE06 functional [6, 7, 8].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
	def2-SV		
$Amylose_8$	-4951.3205620620	4655	2384
DNA_2	-4500.6327971215	5881	2956
$(H_2O)_{68}$	-5185.6849187408	2718	1295
$Polyethyne_{128}$	-4944.6167451692	2017	1115
$(S_8)_5$	-15916.2919545647	350	205
	def2-SVPD		
$Amylose_8$	-4953.5300284379	170481	119923
DNA_2	-4502.5962276056	168659	112484
$(H_2O)_{68}$	-5189.1364561538	203492	126269
$Polyethyne_{128}$	-4946.1647073198	41643	34225
$(S_8)_5$	-15918.3329759315	21479	12172
	def2-TZVP		
$Amylose_8$	-4958.7185624054	128353	97492
DNA_2	-4506.8696669914	153454	114987
$(H_2O)_{68}$	-5194.7265670686	28593	20829
$Polyethyne_{128}$	-4951.2009579457	79660	70270
$(S_8)_5$	-15923.7893787900	11811	9089

Table 7: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC LinK (Eq. 7) calculations on CPUs using the HSE06 functional [6, 7, 8].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
	def2-SV		
$Amylose_8$	-4951.3205620606	3718	1444
DNA_2	-4500.6327971214	4849	1917
$(H_2O)_{68}$	-5185.6849187407	2122	703
$Polyethyne_{128}$	-4944.6167451701	1309	403
$(S_8)_5$	-15916.2919545672	286	133
	def2-SVPD		
$Amylose_8$	-4953.5300284344	115553	64988
DNA_2	-4502.5962276037	132424	76334
$(H_2O)_{68}$	-5189.1364561528	151883	74864
$Polyethyne_{128}$	-4946.1647073130	16312	8876
$(S_8)_5$	-15918.3329759342	18471	9196
	def2-TZVP		
$Amylose_8$	-4958.7185624052	78514	47798
DNA_2	-4506.8696669913	104839	66143
$(H_2O)_{68}$	-5194.7265670680	17182	9430
$Polyethyne_{128}$	-4951.2009579434	25701	16324
$(S_8)_5$	-15923.7893787894	8243	5526

3.1.4 Data for Section 'Different ω-Screening Parameters'

Table 8: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on CPUs using the HSE12s functional [10]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_8$	-4952.7779232697	9466	5568
DNA_2	-4502.1196855058	12076	6975
$(H_2O)_{68}$	-5187.9190641817	5222	2958
$Polyethyne_{128}$	-4945.9069482660	5169	3552
$Polyyne_{1024}$	-38911.7852933736	151584	35361
$(S_8)_5$	-15918.6065844305	688	441
$(S_8)_{20}$	-63674.4940161102	12327	9986

Table 9: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC LinK (Eq. 7) calculations on CPUs using the HSE12s functional [10]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_8$	-4952.7779232702	6372	2481
DNA_2	-4502.1196855057	8352	3244
$(H_2O)_{68}$	-5187.9190641816	3307	1021
$Polyethyne_{128}$	-4945.9069482662	2416	798
$Polyyne_{1024}$	-38911.7852933792	119731	3724
$(S_8)_5$	-15918.6065844316	439	195
$(S_8)_{20}$	-63674.4940161162	4511	2174

Table 10: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on CPUs using the HSE06 functional with varying ω-values. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$(H_2O)_{68}$ $\omega = 0.03$	-5188.2753456259	5277	2994
$(H_2O)_{68}$ $\omega = 0.06$	-5188.2644453995	5252	2998
$(H_2O)_{68}$ $\omega = 0.09$	-5188.2418515808	5236	2972
$(H_2O)_{68}$ $\omega = 0.12$	-5188.2080074242	5244	2974
$(H_2O)_{68}$ $\omega = 0.15$	-5188.1646813287	5341	3025
$(H_2O)_{68}$ $\omega = 0.18$	-5188.1140941219	5301	3018
$(H_2O)_{68}$ $\omega = 0.21$	-5188.0585007055	5315	3026
$(H_2O)_{68}$ $\omega = 0.24$	-5187.9999889148	5330	3037
$(H_2O)_{68}$ $\omega = 0.27$	-5187.9403905412	5307	3022
$(H_2O)_{68}$ $\omega = 0.30$	-5187.8812574021	5229	2990
$(H_2O)_{68}$ $\omega = 0.40$	-5187.7020724959	5295	3030
$(H_2O)_{68}$ $\omega = 0.50$	-5187.5696243854	5307	3042
$(H_2O)_{68}$ $\omega = 0.60$	-5187.4933846844	5324	3047
$(H_2O)_{68}$ $\omega = 0.70$	-5187.4706952243	5489	3215
$(H_2O)_{68}$ $\omega = 0.80$	-5187.4923965782	5319	3033
$(H_2O)_{68}$ $\omega = 0.90$	-5187.5468103389	5329	3051

Table 11: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC LinK (Eq. 7) calculations on CPUs using the HSE06 functional with varying ω-values. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$(H_2O)_{68}$ $\omega = 0.03$	-5188.2753456254	4323	2038
$(H_2O)_{68}$ $\omega = 0.06$	-5188.2644453990	4085	1835
$(H_2O)_{68}$ $\omega = 0.09$	-5188.2418515804	3897	1642
$(H_2O)_{68}$ $\omega = 0.12$	-5188.2080074239	3746	1470
$(H_2O)_{68}$ $\omega = 0.15$	-5188.1646813285	3608	1310
$(H_2O)_{68}$ $\omega = 0.18$	-5188.1140941217	3469	1175
$(H_2O)_{68}$ $\omega = 0.21$	-5188.0585007055	3352	1067
$(H_2O)_{68}$ $\omega = 0.24$	-5187.9999889148	3253	973
$(H_2O)_{68}$ $\omega = 0.27$	-5187.9403905412	3212	902
$(H_2O)_{68}$ $\omega = 0.30$	-5187.8812574022	3112	833
$(H_2O)_{68}$ $\omega = 0.40$	-5187.7020724960	2944	682
$(H_2O)_{68}$ $\omega = 0.50$	-5187.5696243854	2863	590
$(H_2O)_{68}$ $\omega = 0.60$	-5187.4933846844	2812	539
$(H_2O)_{68}$ $\omega = 0.70$	-5187.4706952242	2767	496
$(H_2O)_{68}$ $\omega = 0.80$	-5187.4923965778	2747	470
$(H_2O)_{68}$ $\omega = 0.90$	-5187.5468103372	2726	447

3.2 GPU calculations

3.2.1 Data for Section 'General Purpose Test using the HSE06 functional'

Table 12: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on GPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_{2}$	-1295.4567122199	162	88
$Amylose_{4}$	-2514.6590442432	552	371
$Amylose_{8}$	-4953.0635614375	1480	1066
$Amylose_{16}$	-9829.8723960829	3567	2521
$Amylose_{32}$	-19583.4903947622	7941	5141
$Amylose_{48}$	-29337.1094494225	14453	8596
$Amylose_{64}$	-39090.7291635501	22038	11715
Beta-carotene	-1555.0156201547	495	310
CNT^{*}_{20}	-766.3696902044	95	50
CNT^{*}_{40}	-1526.9882091126	451	317
CNT^{*}_{80}	-3048.2929922719	1786	1438
CNT^{*}_{160}	-6090.8435865208	10345	8977
$CNT(6,3)^{*}_{8}$	-25576.2260633609	98243	86368
$Diamond_{102}$	-1631.7173289979	2070	1522
DNA_{1}	-1760.3240502851	283	170
DNA_{2}	-4502.3006370998	1794	1382
DNA_{4}	-9986.2377487066	7670	6522
DNA_{8}	-20954.1262303313	25187	21818
DNA_{16}	-42889.9083605285	64162	53332
$Graphite_{24}$	-920.2459652908	145	79
$Graphite_{54}$	-2064.5666930582	881	640
$Graphite_{96}$	-3665.9943526149	3068	2487
$(H_2O)_{68}$	-5188.2204457092	909	631
$(H_2O)_{142}$	-10834.5826332639	3553	2732
$(H_2O)_{285}$	-21746.0223241400	16814	13487
$(H_2O)_{569}$	-43416.3055035228	45770	35158
$(LiF)_{32}$	-1717.1001635039	81	40
$(LiF)_{72}$	-3863.8868803791	936	690
$Polyethyne_{64}$	-2473.5765330623	434	227
$Polyethyne_{128}$	-4945.9875093391	1015	485
$Polyyne_{64}$	-2433.2728418121	192	96
$Polyyne_{1024}$	-38914.7735887543	20356	3755
$(S_8)_5$	-15918.0655864665	164	115
$(S_8)_{20}$	-63672.3245813516	3264	2825
DGCR8 Dimerization Domain	-24313.7175975892	50558	42713

Table 13: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC PreLinK calculations on GPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
Amylose_2	-1295.4567122196	163	86
Amylose_4	-2514.6590442428	504	325
Amylose_8	-4953.0635614368	1239	821
Amylose_{16}	-9829.8723960842	2915	1868
Amylose_{32}	-19583.4903947618	6598	3781
Amylose_{48}	-29337.1094494238	12079	6298
Amylose_{64}	-39090.7291635505	18926	8677
Beta-carotene	-1555.0156201529	380	196
CNT^*_{20}	-766.3696902044	95	52
CNT^*_{40}	-1526.9882091128	450	317
CNT^*_{80}	-3048.2929922708	1745	1396
CNT^*_{160}	-6090.8435865152	7865	6496
$\text{CNT}(6,3)^*_8$	-25576.2260632824	57590	45718
Diamond_{102}	-1631.7173289967	2072	1522
DNA_1	-1760.3240502852	254	143
DNA_2	-4502.3006371000	1540	1124
DNA_4	-9986.2377487066	5847	4703
DNA_8	-20954.1262303294	16812	13430
DNA_{16}	-42889.9083605258	41767	31140
Graphite_{24}	-920.2459652915	145	79
Graphite_{54}	-2064.5666930575	873	631
Graphite_{96}	-3665.9943526094	2883	2299
$(\text{H}_2\text{O})_{68}$	-5188.2204457061	751	475
$(\text{H}_2\text{O})_{142}$	-10834.5826332549	2629	1805
$(\text{H}_2\text{O})_{285}$	-21746.0223241208	11406	8052
$(\text{H}_2\text{O})_{569}$	-43416.3055034693	30533	19787
$(\text{LiF})_{32}$	-1717.1001635039	79	39
$(\text{LiF})_{72}$	-3863.8868803794	918	672
Polyethyne_{64}	-2473.5765330516	328	121
Polyethyne_{128}	-4945.9875093159	781	247
Polyyne_{64}	-2433.2728418120	146	52
Polyyne_{1024}	-38914.7735887538	17917	1810
$(\text{S}_8)_5$	-15918.0655864656	143	93
$(\text{S}_8)_{20}$	-63672.3245813414	2093	1652
DGCR8 Dimerization Domain	-24313.7175975879	32359	24645

Table 14: Total energies, total SCF timings and total exchange matrix formation timings for the $T_{\omega}^{0.20}$-Pre-Screening calculations on GPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_{2}$	-1295.4567122194	161	85
$Amylose_{4}$	-2514.6590442418	459	280
$Amylose_{8}$	-4953.0635614336	1075	661
$Amylose_{16}$	-9829.8723960771	2522	1474
$Amylose_{32}$	-19583.4903947535	5784	2967
$Amylose_{48}$	-29337.1094494084	10827	4991
$Amylose_{64}$	-39090.7291635300	17273	6923
Beta-carotene	-1555.0156190315	339	148
CNT_{20}^{*}	-766.3696902044	95	51
CNT_{40}^{*}	-1526.9882088129	449	314
CNT_{80}^{*}	-3048.2929891556	1475	1128
CNT_{160}^{*}	-6090.8435732430	5690	4326
$CNT(6,3)_{8}^{*}$	-25576.2260231392	38641	26782
$Diamond_{102}$	-1631.7173289973	2070	1521
DNA_{1}	-1760.3240502799	234	121
DNA_{2}	-4502.3006370856	1344	928
DNA_{4}	-9986.2377486715	4580	3437
DNA_{8}	-20954.1262302432	12395	9019
DNA_{16}	-42889.9083603671	31280	20455
$Graphite_{24}$	-920.2459652774	144	77
$Graphite_{54}$	-2064.5666887212	779	534
$Graphite_{96}$	-3665.9943441274	2157	1572
$(H_2O)_{68}$	-5188.2204457069	701	421
$(H_2O)_{142}$	-10834.5826332627	2345	1484
$(H_2O)_{285}$	-21746.0223241332	9619	6229
$(H_2O)_{569}$	-43416.3055034871	25014	14614
$(LiF)_{32}$	-1717.1001633764	75	35
$(LiF)_{72}$	-3863.8868808114	829	581
$Polyethyne_{64}$	-2473.5765323427	295	89
$Polyethyne_{128}$	-4945.9875078299	715	183
$Polyyne_{64}$	-2433.2728385219	126	31
$Polyyne_{1024}$	-38914.7735326321	17920	1565
$(S_8)_5$	-15918.0655863282	129	80
$(S_8)_{20}$	-63672.3245803789	1629	1186
DGCR8 Dimerization Domain	-24313.7175974982	23258	15417

3.2.2 Data for Section 'Scaling Behavior'

Table 15: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on GPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Alkane_{01}$	-40.4316067051	2	1
$Alkane_{02}$	-79.6689867315	4	2
$Alkane_{03}$	-118.9088500435	8	0
$Alkane_{04}$	-158.1482944926	11	2
$Alkane_{05}$	-197.3876326755	16	4
$Alkane_{06}$	-236.6269683566	21	4
$Alkane_{07}$	-275.8663020247	29	8
$Alkane_{08}$	-315.1056240516	32	8
$Alkane_{09}$	-354.3449491990	41	12
$Alkane_{10}$	-393.5842688390	48	14
$Alkane_{13}$	-511.3022318665	70	26
$Alkane_{15}$	-589.7808723186	87	30
$Alkane_{20}$	-785.9774725166	125	55
$Alkane_{30}$	-1178.3706725744	225	108
$Alkane_{40}$	-1570.7638719227	334	162
$Alkane_{50}$	-1963.1570713362	444	223
$Alkane_{60}$	-2355.5502709667	558	281
$Alkane_{80}$	-3140.3366702994	813	398
$Alkane_{100}$	-3925.1230696454	1092	516
$Alkane_{160}$	-6279.4822669559	2078	880
$Alkane_{200}$	-7849.0550653718	2990	1136
$Alkane_{300}$	-11772.9870609808	5901	1822
$Alkane_{320}$	-12557.7734602370	6758	1960

Table 16: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC PreLinK calculations on GPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Alkane_{01}$	-40.4316067051	2	≤ 1
$Alkane_{02}$	-79.6689867315	4	≤ 1
$Alkane_{03}$	-118.9088500435	7	1
$Alkane_{04}$	-158.1482944926	11	2
$Alkane_{05}$	-197.3876326755	16	3
$Alkane_{06}$	-236.6269683565	21	4
$Alkane_{07}$	-275.8663020246	28	8
$Alkane_{08}$	-315.1056240515	32	8
$Alkane_{09}$	-354.3449491988	41	11
$Alkane_{10}$	-393.5842688388	47	13
$Alkane_{13}$	-511.3022318670	69	23
$Alkane_{15}$	-589.7808723187	84	30
$Alkane_{20}$	-785.9774725166	115	43
$Alkane_{30}$	-1178.3706725750	192	76
$Alkane_{40}$	-1570.7638719233	275	109
$Alkane_{50}$	-1963.1570713352	362	142
$Alkane_{60}$	-2355.5502709663	454	179
$Alkane_{80}$	-3140.3366702994	661	243
$Alkane_{100}$	-3925.1230696453	894	312
$Alkane_{160}$	-6279.4822669558	1741	542
$Alkane_{200}$	-7849.0550653716	2555	698
$Alkane_{300}$	-11772.9870609811	5227	1152
$Alkane_{320}$	-12557.7734602369	6056	1255

Table 17: Total energies, total SCF timings and total exchange matrix formation timings for the T_ω-Pre-Screening calculations on GPUs using the HSE06 functional [6, 7, 8]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Alkane_{01}$	-40.4316067051	2	0
$Alkane_{02}$	-79.6689867315	4	0
$Alkane_{03}$	-118.9088500435	8	1
$Alkane_{04}$	-158.1482944926	11	1
$Alkane_{05}$	-197.3876326755	16	4
$Alkane_{06}$	-236.6269683565	21	4
$Alkane_{07}$	-275.8663020245	28	3
$Alkane_{08}$	-315.1056240504	31	8
$Alkane_{09}$	-354.3449491961	40	11
$Alkane_{10}$	-393.5842688344	47	14
$Alkane_{13}$	-511.3022318580	67	20
$Alkane_{15}$	-589.7808723073	80	24
$Alkane_{20}$	-785.9774724983	106	33
$Alkane_{30}$	-1178.3706725422	173	58
$Alkane_{40}$	-1570.7638718776	248	86
$Alkane_{50}$	-1963.1570712762	327	110
$Alkane_{60}$	-2355.5502708933	410	132
$Alkane_{80}$	-3140.3366701996	600	181
$Alkane_{100}$	-3925.1230695180	811	235
$Alkane_{160}$	-6279.4822667469	1610	411
$Alkane_{200}$	-7849.0550651083	2394	535
$Alkane_{300}$	-11772.9870605809	4994	901
$Alkane_{320}$	-12557.7734598097	5788	994

3.2.3 Data for Section 'Basis Set Dependence'

Table 18: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on GPUs using the HSE06 functional [6, 7, 8].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
def2-SV			
$Amylose_8$	-4951.3205617632	681	397
DNA_2	-4500.6327972428	801	508
$(H_2O)_{68}$	-5185.6849199231	428	216
$Polyethyne_{128}$	-4944.6167451575	469	142
$(S_8)_5$	-15916.2919544447	68	34
def2-SVPD			
$Amylose_8$	-4953.5300285246	19450	16937
DNA_2	-4502.5962277045	25695	22491
$(H_2O)_{68}$	-5189.1364559127	20224	17476
$Polyethyne_{128}$	-4946.1647089930	4362	2873
$(S_8)_5$	-15918.3329758399	2311	1959
def2-TZVP			
$Amylose_8$	-4958.7185627018	19174	17337
DNA_2	-4506.8696671336	21856	19994
$(H_2O)_{68}$	-5194.7265666859	4461	3919
$Polyethyne_{128}$	-4951.2009598450	9330	7025
$(S_8)_5$	-15923.7893800571	2583	2368

Table 19: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC PreLinK calculations on GPUs using the HSE06 functional [6, 7, 8].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
	def2-SV		
$Amylose_8$	-4951.3205617635	596	311
DNA_2	-4500.6327972428	715	423
$(H_2O)_{68}$	-5185.6849199200	385	171
$Polyethyne_{128}$	-4944.6167451350	405	77
$(S_8)_5$	-15916.2919544440	62	27
	def2-SVPD		
$Amylose_8$	-4953.5300285276	16797	14282
DNA_2	-4502.5962277030	25010	21811
$(H_2O)_{68}$	-5189.1364559145	19880	17123
$Polyethyne_{128}$	-4946.1647089924	3276	1789
$(S_8)_5$	-15918.3329758390	2262	1909
	def2-TZVP		
$Amylose_8$	-4958.7185627021	13692	11862
DNA_2	-4506.8696671328	19286	17428
$(H_2O)_{68}$	-5194.7265666816	3362	2816
$Polyethyne_{128}$	-4951.2009598341	5612	3301
$(S_8)_5$	-15923.7893800539	2326	2112

Table 20: Total energies, total SCF timings and total exchange matrix formation timings for the $T_\omega^{0.20}$-Pre-Screening calculations on GPUs using the HSE06 functional [6, 7, 8].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
	def2-SV		
$Amylose_8$	-4951.3205617600	531	247
DNA_2	-4500.6327972319	640	344
$(H_2O)_{68}$	-5185.6849199213	365	153
$Polyethyne_{128}$	-4944.6167448189	383	56
$(S_8)_5$	-15916.2919543091	59	23
	def2-SVPD		
$Amylose_8$	-4953.5300285778	10905	8394
DNA_2	-4502.5962276934	18976	15772
$(H_2O)_{68}$	-5189.1364559071	15332	12578
$Polyethyne_{128}$	-4946.1647083712	2502	1020
$(S_8)_5$	-15918.3329759660	1925	1573
	def2-TZVP		
$Amylose_8$	-4958.7185626792	9925	8096
DNA_2	-4506.8696671135	14945	13088
$(H_2O)_{68}$	-5194.7265666805	2877	2339
$Polyethyne_{128}$	-4951.2009483657	4446	2142
$(S_8)_5$	-15923.7893799622	1862	1648

3.2.4 Data for Section 'Different ω-Screening Parameters'

Table 21: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on GPUs using the HSE12s functional [10]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_8$	-4952.7779231156	1541	1135
DNA_2	-4502.1196856617	1878	1472
$(H_2O)_{68}$	-5187.9190652300	991	712
$Polyethyne_{128}$	-4945.9069491412	830	365
$Polyyne_{1024}$	-38911.7852953040	24202	3346
$(S_8)_5$	-15918.6065843363	167	121
$(S_8)_{20}$	-63674.4940131777	3253	2849

Table 22: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC PreLinK calculations on GPUs using the HSE12s functional [10]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_8$	-4952.7779231154	1241	830
DNA_2	-4502.1196856622	1551	1144
$(H_2O)_{68}$	-5187.9190652302	779	497
$Polyethyne_{128}$	-4945.9069491412	696	227
$Polyyne_{1024}$	-38911.7852953041	22648	1724
$(S_8)_5$	-15918.6065843363	134	90
$(S_8)_{20}$	-63674.4940131780	1929	1525

Table 23: Total energies, total SCF timings and total exchange matrix formation timings for the $T_\omega^{0.20}$-Pre-Screening calculations on GPUs using the HSE12s functional [10]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_8$	-4952.7779209267	833	428
DNA_2	-4502.1196846049	968	561
$(H_2O)_{68}$	-5187.9190633707	482	206
$Polyethyne_{128}$	-4945.9069886016	609	134
$(S_8)_5$	-15918.6065840940	88	45
$(S_8)_{20}$	-63674.4940234064	879	471

We want to note that the test calculations for the $T_\omega^{0.25}$- and $T_\omega^{0.30}$-Pre-Screening caluclations were performed seperately and therefore the reference calculations were repeated on the exact computer node used for the relevant calculation, to obtain precise timings. Therefore the reference values used to generate the numbers in the main article for the $T_\omega^{0.25}$- and $T_\omega^{0.30}$-Pre-Screening caluclations might differ slightly from the reference results shown here, which were used to compare the ERFC PreLinK and $T_\omega^{0.20}$-Pre-Screening caluclations.

Table 24: Total energies, total SCF timings and total exchange matrix formation timings for the $T_\omega^{0.25}$-Pre-Screening calculations on GPUs using the HSE12s functional [10]. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$Amylose_8$	-4952.7779231182	968	561
DNA_2	-4502.1196856472	1178	773
$(H_2O)_{68}$	-5187.9190652223	590	312
$Polyethyne_{128}$	-4945.9069486642	684	161
$Polyyne_{1024}$	-38911.7852809211	25143	1631
$(S_8)_5$	-15918.6065843638	109	62
$(S_8)_{20}$	-63674.4940133571	1176	736

Table 25: Total energies, total SCF timings and total exchange matrix formation timings for the reference calculations on GPUs using the HSE06 functional with ω-values. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$(H_2O)_{68}$ $\omega = 0.03$	-5188.2753466264	757	482
$(H_2O)_{68}$ $\omega = 0.06$	-5188.2644464002	823	547
$(H_2O)_{68}$ $\omega = 0.09$	-5188.2418525830	838	589
$(H_2O)_{68}$ $\omega = 0.12$	-5188.2080084065	885	631
$(H_2O)_{68}$ $\omega = 0.15$	-5188.1646823155	920	670
$(H_2O)_{68}$ $\omega = 0.18$	-5188.1140951214	981	708
$(H_2O)_{68}$ $\omega = 0.21$	-5188.0585017210	964	714
$(H_2O)_{68}$ $\omega = 0.24$	-5187.9999899231	1003	732
$(H_2O)_{68}$ $\omega = 0.27$	-5187.9403915559	988	714
$(H_2O)_{68}$ $\omega = 0.30$	-5187.8812584272	959	687
$(H_2O)_{68}$ $\omega = 0.40$	-5187.7020735197	855	581
$(H_2O)_{68}$ $\omega = 0.50$	-5187.5696253795	776	503
$(H_2O)_{68}$ $\omega = 0.60$	-5187.4933857073	728	455
$(H_2O)_{68}$ $\omega = 0.70$	-5187.4706962333	693	421
$(H_2O)_{68}$ $\omega = 0.80$	-5187.4923975805	675	400
$(H_2O)_{68}$ $\omega = 0.90$	-5187.5468113194	653	380

Table 26: Total energies, total SCF timings and total exchange matrix formation timings for the ERFC PreLinK calculations on GPUs using the HSE06 functional with ω-values. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$(H_2O)_{68}$ $\omega = 0.03$	-5188.2753466038	667	389
$(H_2O)_{68}$ $\omega = 0.06$	-5188.2644463850	708	432
$(H_2O)_{68}$ $\omega = 0.09$	-5188.2418525764	702	450
$(H_2O)_{68}$ $\omega = 0.12$	-5188.2080084041	724	471
$(H_2O)_{68}$ $\omega = 0.15$	-5188.1646823150	733	484
$(H_2O)_{68}$ $\omega = 0.18$	-5188.1140951217	775	504
$(H_2O)_{68}$ $\omega = 0.21$	-5188.0585017210	750	499
$(H_2O)_{68}$ $\omega = 0.24$	-5187.9999899221	782	509
$(H_2O)_{68}$ $\omega = 0.27$	-5187.9403915555	774	502
$(H_2O)_{68}$ $\omega = 0.30$	-5187.8812584278	765	490
$(H_2O)_{68}$ $\omega = 0.40$	-5187.7020735196	710	435
$(H_2O)_{68}$ $\omega = 0.50$	-5187.5696253791	661	386
$(H_2O)_{68}$ $\omega = 0.60$	-5187.4933857083	623	348
$(H_2O)_{68}$ $\omega = 0.70$	-5187.4706962338	593	320
$(H_2O)_{68}$ $\omega = 0.80$	-5187.4923975809	576	301
$(H_2O)_{68}$ $\omega = 0.90$	-5187.5468113190	559	283

Table 27: Total energies, total SCF timings and total exchange matrix formation timings for the $T_{\omega}^{0.20}$-Pre-Screening calculations on GPUs using the HSE06 functional with varying ω-values. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$(H_2O)_{68}$ $\omega = 0.03$	-5188.2753466260	758	485
$(H_2O)_{68}$ $\omega = 0.06$	-5188.2644464004	823	549
$(H_2O)_{68}$ $\omega = 0.09$	-5188.2418525829	745	494
$(H_2O)_{68}$ $\omega = 0.12$	-5188.2080084039	629	379
$(H_2O)_{68}$ $\omega = 0.15$	-5188.1646822912	549	299
$(H_2O)_{68}$ $\omega = 0.18$	-5188.1140945680	527	252
$(H_2O)_{68}$ $\omega = 0.21$	-5188.0585007619	463	210
$(H_2O)_{68}$ $\omega = 0.24$	-5187.9999877066	465	189
$(H_2O)_{68}$ $\omega = 0.27$	-5187.9403873364	444	168
$(H_2O)_{68}$ $\omega = 0.30$	-5187.8812579206	428	156
$(H_2O)_{68}$ $\omega = 0.40$	-5187.7021316552	397	120
$(H_2O)_{68}$ $\omega = 0.50$	-5187.5697943203	376	101
$(H_2O)_{68}$ $\omega = 0.60$	-5187.4936410063	361	88
$(H_2O)_{68}$ $\omega = 0.70$	-5187.4709473905	356	83
$(H_2O)_{68}$ $\omega = 0.80$	-5187.4926412020	351	78
$(H_2O)_{68}$ $\omega = 0.90$	-5187.5469893517	348	75

Table 28: Total energies, total SCF timings and total exchange matrix formation timings for the $T_{\omega}^{0.25}$-Pre-Screening calculations on GPUs using the HSE06 functional with varying ω-values. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$(H_2O)_{68}$ $\omega = 0.03$	-5188.2753466258	713	486
$(H_2O)_{68}$ $\omega = 0.06$	-5188.2644464010	779	552
$(H_2O)_{68}$ $\omega = 0.09$	-5188.2418525827	825	595
$(H_2O)_{68}$ $\omega = 0.12$	-5188.2080084064	806	532
$(H_2O)_{68}$ $\omega = 0.15$	-5188.1646823157	718	443
$(H_2O)_{68}$ $\omega = 0.18$	-5188.1140951221	649	374
$(H_2O)_{68}$ $\omega = 0.21$	-5188.0585017131	604	326
$(H_2O)_{68}$ $\omega = 0.24$	-5187.9999899067	561	285
$(H_2O)_{68}$ $\omega = 0.27$	-5187.9403915072	535	261
$(H_2O)_{68}$ $\omega = 0.30$	-5187.8812583458	513	238
$(H_2O)_{68}$ $\omega = 0.40$	-5187.7020728276	420	189
$(H_2O)_{68}$ $\omega = 0.50$	-5187.5696243423	391	162
$(H_2O)_{68}$ $\omega = 0.60$	-5187.4933856852	369	141
$(H_2O)_{68}$ $\omega = 0.70$	-5187.4706975861	357	131
$(H_2O)_{68}$ $\omega = 0.80$	-5187.4923991213	350	119
$(H_2O)_{68}$ $\omega = 0.90$	-5187.5468124877	343	115

Table 29: Total energies, total SCF timings and total exchange matrix formation timings for the $T_{\omega}^{0.30}$-Pre-Screening calculations on GPUs using the HSE06 functional with varying ω-values. The calculations were performed using the def2-SVP basis set [9].

System	tot. Energy [H]	tot. SCF Time [s]	tot. Exchange Time[s]
$(H_2O)_{68}$ $\omega = 0.03$	-5188.2753466258	758	483
$(H_2O)_{68}$ $\omega = 0.06$	-5188.2644464009	822	547
$(H_2O)_{68}$ $\omega = 0.09$	-5188.2418525834	871	599
$(H_2O)_{68}$ $\omega = 0.12$	-5188.2080084063	890	617
$(H_2O)_{68}$ $\omega = 0.15$	-5188.1646823156	830	557
$(H_2O)_{68}$ $\omega = 0.18$	-5188.1140951213	763	491
$(H_2O)_{68}$ $\omega = 0.21$	-5188.0585017208	707	435
$(H_2O)_{68}$ $\omega = 0.24$	-5187.9999899223	639	390
$(H_2O)_{68}$ $\omega = 0.27$	-5187.9403915554	630	356
$(H_2O)_{68}$ $\omega = 0.30$	-5187.8812584273	598	324
$(H_2O)_{68}$ $\omega = 0.40$	-5187.7020735171	537	257
$(H_2O)_{68}$ $\omega = 0.50$	-5187.5696253685	494	220
$(H_2O)_{68}$ $\omega = 0.60$	-5187.4933856741	470	193
$(H_2O)_{68}$ $\omega = 0.70$	-5187.4706961696	449	176
$(H_2O)_{68}$ $\omega = 0.80$	-5187.4923975037	438	164
$(H_2O)_{68}$ $\omega = 0.90$	-5187.5468112081	429	157

References

[1] Christian Ochsenfeld, Christopher A. White, and Martin Head-Gordon. Linear and sublinear scaling formation of hartreefock-type exchange matrices. The Journal of Chemical Physics, 109(5):1663–1669, 1998.

[2] Simon A. Maurer, Daniel S. Lambrecht, Denis Flaig, and Christian Ochsenfeld. Distance-dependent schwarz-based integral estimates for two-electron integrals: Reliable tightness vs. rigorous upper bounds. The Journal of Chemical Physics, 136(14), 2012.

[3] See http://www.cup.uni-muenchen.de/pc/ochsenfeld/download.html for structure files.

[4] Rachel Senturia, Arthur Laganowsky, Ian Barr, Brooke D. Scheidemantle, and Feng Guo. Dimerization and heme binding are conserved in amphibian and starfish homologues of the microrna processing protein dgcr8. PLOS ONE, 7(7):1–8, 07 2012.

[5] Marcus D. Hanwell, Donald E. Curtis, David C. Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics, 4(1):17, 2012.

[6] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. The Journal of Chemical Physics, 118(18):8207–8215, 2003.

[7] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Erratum: hybrid functionals based on a screened coulomb potential [j. chem. phys.118, 8207 (2003)]. The Journal of Chemical Physics, 124(21), 2006.

[8] Aliaksandr V. Krukau, Oleg A. Vydrov, Artur F. Izmaylov, and Gustavo E. Scuseria. Influence of the exchange screening parameter on the performance of screened hybrid functionals. The Journal of Chemical Physics, 125(22), 2006.

[9] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. Phys. Chem. Chem. Phys., 7:3297–3305, 2005.

[10] Jonathan E. Moussa, Peter A. Schultz, and James R. Chelikowsky. Analysis of the heyd-scuseria-ernzerhof density functional parameter space. The Journal of Chemical Physics, 136(20), 2012.

3.2 Publication II: Accurate and Efficient Parallel Implementation of an Effective Linear-Scaling Direct Random Phase Approximation Method

D. Graf, M. Beuerle, H. F. Schurkus, A. Luenser, G. Savasci, C. Ochsenfeld,
"Accurate and Efficient Parallel Implementation of an Effective Linear-Scaling Direct Random Phase Approximation Method",
J. Chem. Theory Comput., **14**, 2505 (2018).

Abstract: An efficient algorithm for calculating the random phase approximation (RPA) correlation energy is presented that is as accurate as the canonical molecular orbital resolution-of-the-identity RPA (RI-RPA) with the important advantage of an effective linear-scaling behavior (instead of quartic) for large systems due to a formulation in the local atomic orbital space. The high accuracy is achieved by utilizing optimized minimax integration schemes and the local Coulomb metric attenuated by the complementary error function for the RI approximation. The memory bottleneck of former atomic orbital (AO)-RI-RPA implementations (Schurkus, H. F.; Ochsenfeld, C. *J. Chem. Phys.* **2016**, 144, 031101 and Luenser, A.; Schurkus, H. F.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2017**, 13, 1647–1655) is addressed by precontraction of the large 3-center integral matrix with the Cholesky factors of the ground state density reducing the memory requirements of that matrix by a factor of $\frac{N_{basis}}{N_{occ}}$. Furthermore, we present a parallel implementation of our method, which not only leads to faster RPA correlation energy calculations but also to a scalable decrease in memory requirements, opening the door for investigations of large molecules even on small- to medium-sized computing clusters. Although it is known that AO methods are highly efficient for extended systems, where sparsity allows for reaching the linear-scaling regime, we show that our work also extends the applicability when considering highly delocalized systems for which no linear scaling can be achieved. As an example, the interlayer distance of two covalent organic framework pore fragments (comprising 384 atoms in total) is analyzed.

JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 2505−2515

pubs.acs.org/JCTC

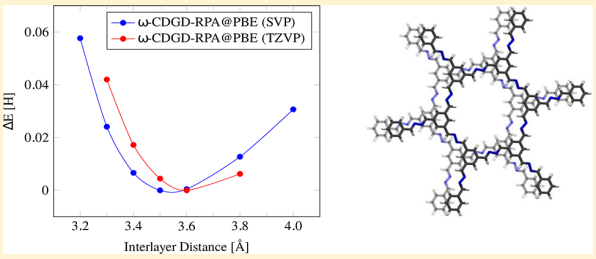
Accurate and Efficient Parallel Implementation of an Effective Linear-Scaling Direct Random Phase Approximation Method

Daniel Graf,[†] Matthias Beuerle,[†] Henry F. Schurkus,[†] Arne Luenser,[†] Gökcen Savasci,[‡,†] and Christian Ochsenfeld*[,†,‡]

[†]Chair of Theoretical Chemistry and Center for Integrated Protein Science Munich (CIPSM), Department of Chemistry, University of Munich (LMU), D-81377 Munich, Germany

[‡]Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany

ABSTRACT: An efficient algorithm for calculating the random phase approximation (RPA) correlation energy is presented that is as accurate as the canonical molecular orbital resolution-of-the-identity RPA (RI-RPA) with the important advantage of an effective linear-scaling behavior (instead of quartic) for large systems due to a formulation in the local atomic orbital space. The high accuracy is achieved by utilizing optimized minimax integration schemes and the local Coulomb metric attenuated by the complementary error function for the RI approximation. The memory bottleneck of former atomic orbital (AO)-RI-RPA implementations (Schurkus, H. F.; Ochsenfeld, C. *J. Chem. Phys.* **2016**, *144*, 031101 and Luenser, A.; Schurkus, H. F.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2017**, *13*, 1647−1655) is addressed by precontraction of the large 3-center integral matrix with the Cholesky factors of the ground state density reducing the memory requirements of that matrix by a factor of $\frac{N_{basis}}{N_{occ}}$. Furthermore, we present a parallel implementation of our method, which not only leads to faster RPA correlation energy calculations but also to a scalable decrease in memory requirements, opening the door for investigations of large molecules even on small- to medium-sized computing clusters. Although it is known that AO methods are highly efficient for extended systems, where sparsity allows for reaching the linear-scaling regime, we show that our work also extends the applicability when considering highly delocalized systems for which no linear scaling can be achieved. As an example, the interlayer distance of two covalent organic framework pore fragments (comprising 384 atoms in total) is analyzed.



1. INTRODUCTION

Density functional theory (DFT) is the most widely used electronic structure method in chemistry, physics, and materials sciences. This is mainly because of the excellent cost-performance ratios and good accuracies of its parametrized functionals for certain types of compounds and properties. However, the high sensitivity of semilocal DFT results to the parametrization has led to the development of many hundreds of different functionals, which not only makes the selection of a suitable functional for a specific problem challenging[1] but also limits its predictive power. Additionally, the general failure of GGA functionals in describing noncovalent interactions[2,3] necessitates the development of more broadly applicable correlation models.

One theory to describe electron correlation, which has become increasingly popular over the past decade, is the random phase approximation (RPA). RPA is a post Kohn−Sham[4] method that was originally introduced by Bohm and Pines in 1953.[5] It contains an ab initio description of dispersion effects,[6] is size consistent,[7] and does not depend on any empirical parameters. Additionally, it is applicable to vanishing electronic gap systems,[7−9] making it highly interesting for a wide range of applications in the field of quantum chemistry.

However, in its original form,[10,11] the calculation of RPA correlation energies of molecular systems scales as $O(M^6)$ with system size M, making it impractical for larger molecules. In 2010, Furche and co-workers[12−14] successfully employed the resolution-of-the-identity (RI) approximation to reduce the scaling to effective $O(M^4)$, which can be viewed as a breakthrough because it opened the door for RPA calculations beyond the few atoms scale. In 2014, Kresse and co-workers[15] presented an algorithm for the RPA that employed minimax grids for the time as well as the frequency domain and a Fourier transform scheme to reduce the scaling to cubic. Effective linear-scaling of RPA calculations for molecules with non-vanishing band gaps was achieved by Kállay[16] as well as Schurkus and Ochsenfeld:[17] The approach of Kállay[16] is based on local correlation theory, which finds its origin in the 1980s,[18−20] whereas the method of Schurkus et al.[17] builds upon the transformation into the local atomic orbital space. Recently, Hutter and co-workers[21] presented a different approach toward large-scale RPA calculations by a cubic scaling but highly parallel implementation, opening the door to

Received: February 17, 2018
Published: April 16, 2018

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

calculations comprising up to 8000 electrons on a Cray XC40 supercomputer.

To obtain the desired linear-scaling behavior, Schurkus et al.[17] employed the local but with canonical auxiliary basis sets less accurate overlap RI metric[22] instead of the longe-range Coulomb metric. Additionally, the pilot implementation was strongly hampered in its applicability to larger basis sets because the scaling with basis set size N_{basis} and auxiliary basis set size N_{aux} is increased from $O(N_{aux}^2 N_{basis} N_{occ})$ to $O(N_{aux}^2 N_{basis}^2)$ for a fixed molecular size. In 2017, Luenser et al.[23] remedied these drawbacks by switching to the Coulomb metric attenuated by the complementary error function[24−26] as well as utilizing pivoted Cholesky decomposition[27−31] of density and pseudodensity matrices, which brought the scaling with basis set size back to $O(N_{aux}^2 N_{basis} N_{occ})$ while scaling linearly with molecular size. The improved implementation was termed ω-CDD-RPA.[23]

However, in both the original atomic orbital RI-RPA as well as the ω-CDD method, the complete 3-center integral matrix $\mathbf{B}$ in the atomic orbital (AO) basis needs to be stored in memory, which is a bottleneck when aiming for large systems. Additionally, two different formulations for constructing the frequency-dependent energy kernel $\mathbf{Q}(u)$ are necessary to avoid numerical problems arising from the use of unoptimized weights and roots for the numerical quadratures. Furthermore, the final frequency integration is carried out with the Clenshaw−Curtis scheme[32] with optimization of the scaling parameter as described by Furche and co-workers,[13] where at least 60 node points are necessary to obtain μHartree accuracy.

The focus of the present work is to overcome the drawbacks described above: The use of an optimized minimax grid for the frequency integration as described by Kresse and co-workers[15] reduces the number of necessary quadrature points from 60−100 to 10−20. Utilizing optimized weights and roots for the integrated double-Laplace expansion[17] enables us to obtain $\mathbf{Q}(u)$ without using the second equation, decreasing the prefactor of our RPA reformulation by a factor of 4 and, additionally, increasing the accuracy of the calculations by up to 4 orders of magnitude. Reformulating the formation of $\mathbf{F}_{INT}$ as described in this work reduces the memory effort of the algorithm by a factor of $\frac{N_{basis}}{N_{occ}}$ and yields an additional speed-up. Besides improved integration schemes, a parallel implementation of the new method is presented, which not only reduces the evaluation time of correlation energies significantly but also leads to a scalable decrease in memory requirements. Here, the focus of our present parallel implementation is on small- to medium-sized computing clusters typically available in local research groups.

In the following, we first give a brief review of the derivation of our ω-CDD-RI-RPA method[23] in section 2.1. We then establish the connection of our double-Laplace approach[17] and the cosine transformation of Kresse and co-workers[15] in section 2.2. In section 2.3, the memory efficient reformulation of evaluating $\mathbf{F}_{INT}$ is outlined before we present the parallel implementation of the method to further reduce the memory requirements and speed up the calculations in section 2.4. Computational details are given in section 3. Accuracy and performance benchmarks as well as a illustrative application of the new implementation are reported in section 4, followed by the conclusions in section 5.

2. THEORY

2.1. Linear-Scaling Atomic Orbital Random Phase Approximation.

To create a complete picture of our new effective linear-scaling atomic orbital RPA method, we will briefly review the most important steps in the derivation of our recently reported ω-CDD method.[23] In this work, the following notation has been adopted: μ, ν, λ, σ denote atomic orbitals (AOs); i, j denote occupied molecular orbitals (MOs); a, b denote virtual MOs; $\underline{i}$, $\underline{j}$ denote Cholesky orbitals; and M, N denote auxiliary RI functions. The number of basis functions is represented by N_{basis}; the number of auxiliary RI functions is represented by N_{aux}; and the numbers of occupied and virtual molecular orbitals are represented by N_{occ} and N_{virt}, respectively. For 2-, 3-, and 4-center integrals, the Mulliken notation will be employed. Furthermore, Einstein's sum convention[33] is used, and the spin index is dropped for convenience.

Within the adiabatic connection formalism,[34] the total energy can be expressed as[10,11]

$$E = E_T[\{\phi_{KS}\}] + E_J[\{\phi_{KS}\}] + E_X[\{\phi_{KS}\}] + E_C \tag{1}$$

where E_T, E_J, and E_X denote the kinetic, Coulomb, and exact exchange energies, respectively. The expression for the correlation energy[35] E_C, obtained by using the zero-temperature fluctuation−dissipation theorem and the RPA,[36] is given by

$$E_C = \int_{-\infty}^{+\infty} \frac{du}{4\pi} \mathrm{Tr}[\ln(1 - v\chi_0(iu)) + v\chi_0(iu)] \tag{2}$$

where v represents the Coulomb operator

$$v(\mathbf{r}, \mathbf{r}') = \frac{1}{|\mathbf{r} - \mathbf{r}'|} \tag{3}$$

with the electronic coordinates $\mathbf{r}$ and $\mathbf{r}'$, and χ_0 denotes the noninteracting density−density response function in the occupied and virtual orbital representation in the zero-temperature case[37]

$$\chi_0(\mathbf{r}, \mathbf{r}', iu) = -\frac{\phi_i^*(\mathbf{r})\phi_a(\mathbf{r})\phi_a^*(\mathbf{r}')\phi_i(\mathbf{r}')}{\varepsilon_a - \varepsilon_i - iu} - \frac{\phi_i^*(\mathbf{r}')\phi_a(\mathbf{r}')\phi_a^*(\mathbf{r})\phi_i(\mathbf{r})}{\varepsilon_a - \varepsilon_i + iu} \tag{4}$$

with the occupied and virtual molecular spin orbitals ϕ_i and ϕ_a and their respective orbital energies ε_i and ε_a. Note that v and $\chi_0(iu)$ denote matrices of which $v(\mathbf{r}, \mathbf{r}')$ and $\chi_0(\mathbf{r}, \mathbf{r}', iu)$ represent the respective $(\mathbf{r}, \mathbf{r}')$-th element and that

$$\mathrm{Tr}[v\chi_0(iu)] = \int\int \mathbf{dr}\mathbf{dr}' v(\mathbf{r}, \mathbf{r}')\chi_0(\mathbf{r}, \mathbf{r}', iu) \tag{5}$$

The correlation energy can also be expressed in a supermatrix formalism[38,39]

$$E_C = \int_{-\infty}^{+\infty} \frac{du}{4\pi} \mathrm{Tr}[\ln(\mathbf{1} - \mathbf{\Pi}_0(iu)\underline{\underline{\mathbf{V}}}) + \mathbf{\Pi}_0(iu)\underline{\underline{\mathbf{V}}}] \tag{6}$$

where we introduced the Kohn−Sham (KS) polarization propagator in the canonical orbital space

$$\mathbf{\Pi}_0(iu) = \begin{pmatrix} \boldsymbol{\pi}^-(iu) & 0 \\ 0 & \boldsymbol{\pi}^+(iu) \end{pmatrix} \tag{7}$$

with

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

$$\boldsymbol{\pi}^{\pm}(iu) = -(\boldsymbol{\Delta} \pm iu\mathbf{1})^{-1} \tag{8}$$

$$\Delta_{ia,\mathrm{jb}} = \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) \tag{9}$$

and the Hartree kernel matrix

$$\underline{\underline{\mathbf{V}}} = \int\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\mathbf{b}(\mathbf{r})v(\mathbf{r},\mathbf{r}')\mathbf{b}^{\dagger}(\mathbf{r}') = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}^{*} & \mathbf{V}_{11}^{*} \end{pmatrix} \tag{10}$$

with

$$\mathbf{b}^{T}(\mathbf{r}) = (\cdots, \phi_i(\mathbf{r})\phi_a^{*}(\mathbf{r}), \cdots, \phi_i^{*}(\mathbf{r})\phi_a(\mathbf{r}), \cdots) \tag{11}$$

and its submatrices

$$V_{11,iajb} = (ia|jb) \tag{12}$$

$$V_{12,iajb} = (ia|bj) \tag{13}$$

For real-valued orbitals

$$\mathbf{V}_{11} = \mathbf{V}_{11}^{*} = \mathbf{V}_{12} = \mathbf{V}_{12}^{*} \tag{14}$$

holds and, therefore, the Hartree kernel matrix simplifies to

$$\underline{\underline{\mathbf{V}}} = \begin{pmatrix} \mathbf{V} & \mathbf{V} \\ \mathbf{V} & \mathbf{V} \end{pmatrix} \tag{15}$$

The matrices appearing in eq 6 have dimensions ($2N_{\mathrm{PH}} \times 2N_{\mathrm{PH}}$) with N_{PH} denoting the number of products between occupied and virtual orbitals (particle-hole). To reduce the dimensions to ($N_{\mathrm{PH}} \times N_{\mathrm{PH}}$), we use the series expansion of the matrix logarithm

$$\mathrm{Tr}[\ln(\mathbf{1} - \boldsymbol{\Pi}_0(iu)\underline{\underline{\mathbf{V}}})] = -\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}[(\boldsymbol{\Pi}_0(iu)\underline{\underline{\mathbf{V}}})^n] \tag{16}$$

Application of the unitary transformation[40]

$$\mathbf{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & -\mathbf{1} \end{pmatrix} \tag{17}$$

cyclic permutation of the matrix products and considering that only the trace is relevant leads to

$$-\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}[(\mathbf{U}^{\dagger}\boldsymbol{\Pi}_0(iu)\underline{\underline{\mathbf{V}}}\mathbf{U})^n] = -\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}[(\{\boldsymbol{\pi}^{-}(iu) + \boldsymbol{\pi}^{+}(iu)\}\mathbf{V})^n] \tag{18}$$

$$-\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}[(\mathbf{U}^{\dagger}\boldsymbol{\Pi}_0(iu)\underline{\underline{\mathbf{V}}}\mathbf{U})^n] = -\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}[(-2\mathbf{G}(u)\mathbf{V})^n] \tag{19}$$

where

$$\mathbf{G}(u) = \frac{\boldsymbol{\Delta}}{\boldsymbol{\Delta}^2 + u^2\mathbf{1}} \tag{20}$$

Applying the Coulomb-RI metric attenuated by the complementary error function[24,26] as described by Luenser et al.[23]

$$(ia|jb) \approx B_{ia}^{M}\tilde{C}_{MN}B_{jb}^{N} \tag{21}$$

$$B_{ia}^{M} = \left(ia\left|\frac{\mathrm{erfc}(\omega r_{12})}{r_{12}}\right|M\right) \tag{22}$$

$$\tilde{\mathbf{C}} = \mathbf{S}^{-1}\mathbf{C}\mathbf{S}^{-1} \tag{23}$$

$$C_{MN} = (M|r_{12}^{-1}|N) \tag{24}$$

$$S_{MN} = \left(M\left|\frac{\mathrm{erfc}(\omega r_{12})}{r_{12}}\right|N\right) \tag{25}$$

to factorize the 4-center integrals into 3-center integrals B_{ia}^{M}, and 2-center-2-electron integrals $\tilde{C}_{MN}$ leads to

$$-\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}[(-2\mathbf{G}(u)\mathbf{V})^n] = -\sum_{n=1}^{\infty}\frac{1}{n}\mathrm{Tr}[(-2\mathbf{G}(u)\mathbf{B}\tilde{\mathbf{C}}\mathbf{B}^{T})^n] \tag{26}$$

Cyclic permutation of the matrix products and defining

$$\mathbf{Q}(u) = 2\mathbf{B}^{T}\mathbf{G}(u)\mathbf{B} \tag{27}$$

finally yields[17]

$$E_{\mathrm{C}} = 2\int_0^{+\infty}\frac{\mathrm{d}u}{4\pi}\mathrm{Tr}[\ln(\mathbf{1} + \mathbf{Q}(u)\tilde{\mathbf{C}}) - \mathbf{Q}(u)\tilde{\mathbf{C}}] \tag{28}$$

As shown by Schurkus and Ochsenfeld,[17] the Kohn−Sham energy-based $\mathbf{G}(u)$ can be decoupled by a contracted double-Laplace expansion

$$(\mathbf{G}(u))_{ia,jb} = \frac{\Delta_{ia,jb}}{\Delta_{ia,jb}^2 + u^2} \tag{29}$$

$$(\mathbf{G}(u))_{ia,jb} = \Delta_{ia,\mathrm{jb}}\int_0^{\infty}\frac{\sin(up)}{u}\exp(-\Delta_{ia,\mathrm{jb}}p)\,\mathrm{d}p \tag{30}$$

yielding two different equations for calculating $\mathbf{Q}(u)$

$$\mathbf{Q}(u) = 2\int_0^{\infty}\cos(up)\mathbf{F}_{\mathrm{INT}}^{(p)}\,\mathrm{d}p \tag{31}$$

$$\mathbf{Q}(u) = 2\frac{\mathbf{F}^{(0)}}{u^2} + 2\int_0^{\infty}\frac{\cos(up)}{u^2}\mathbf{F}_{\mathrm{D}}^{(p)}\,\mathrm{d}p \tag{32}$$

after partial integration with three different forms of $\mathbf{F}$ matrices defined in ref 23. As evaluating the trace of the matrix logarithm in the final frequency integration (eq 28) has a very small prefactor, the time determining step of the algorithm is the calculation of these three $\mathbf{F}$ matrices. Rewriting the expressions for these matrices in the local atomic orbital basis allows for a linear-scaling calculation because all occurring quantities become sparse for large systems.[17]

2.2. Polarization Propagators in the Imaginary Frequency and Time Domain in AO-RI-RPA Theory. As mentioned in the section above, within the ω-CDD method two equations for the calculation of $\mathbf{Q}(u)$ are necessary because eq 31 suffers from numerical instability when u approaches infinity and eq 32 becomes numerically unstable where u tends to zero.[17] It is assumed that these problems occur because weights and roots are used for the numerical quadratures, which are not optimized for these specific transformations.[17] Therefore, determining optimized weights and roots for the integrated double-Laplace expansion should allow for using eq 31 alone for calculating $\mathbf{Q}(u)$. This would decrease the computational cost significantly while at the same time increasing the accuracy of the quadrature. Kresse and coworkers[15] have described in great detail a procedure for finding optimal weights and roots for their nonuniform cosine transformation, which is based on minimax grids and allows for transforming the polarizability $\hat{\chi}$ from the imaginary time domain into the imaginary frequency domain and vice versa. We show in the following that the described cosine transformation is equivalent to eq 31 stemming from the

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

double-Laplace transform, which enables us to follow the procedure outlined by Kresse and co-workers to obtain optimized weights and roots for this transformation.

Therefore, consider the definition of $\mathbf{Q}(u)$

$$\mathbf{Q}_{MN}(u) = 2B_{ia}^{M}\frac{\varepsilon_a - \varepsilon_i}{u^2 + (\varepsilon_a - \varepsilon_i)^2}B_{ia}^{N} \tag{33}$$

Combination of the two indices i and a according to $\kappa = (i, a)$ and defining

$$x_\kappa = \varepsilon_a - \varepsilon_i \tag{34}$$

results in

$$\mathbf{Q}_{MN}(u) = B_\kappa^M \Phi(u, x_\kappa) B_\kappa^N \tag{35}$$

where

$$\Phi(u, x_\kappa) = \frac{2x_\kappa}{u^2 + x_\kappa^2} \tag{36}$$

Next, consider the definition of $\mathbf{F}_{\mathrm{INT}}$, which is given by[23]

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \mathrm{Tr}(\underline{\mathbf{P}}^{(p)}\mathbf{B}^M\overline{\mathbf{P}}^{(p)}\mathbf{B}^N) \tag{37}$$

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \underline{P}_{\mu\nu}^{(p)}B_{\nu\sigma}^M\overline{P}_{\sigma\lambda}^{(p)}B_{\lambda\mu}^N \tag{38}$$

Inserting the definition for the occupied and virtual pseudodensities, $\underline{P}^{(p)}$ and $\overline{P}^{(p)}$[17] and transforming $\mathbf{B}$ back into the molecular orbital space yields

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = B_{ia}^M \exp(+\varepsilon_i p)\exp(-\varepsilon_a p)B_{ia}^N \tag{39}$$

Considering the definitions for κ and x_κ finally leads to

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = B_\kappa^M \hat{\Phi}(p, x_\kappa)B_\kappa^N \tag{40}$$

where

$$\hat{\Phi}(p, x_\kappa) = \exp(-x_\kappa p) \tag{41}$$

Comparison of eqs 35−40 with eqs 14−17 in ref 15 shows that $\mathbf{F}_{\mathrm{INT}}$ is the representation of $\mathbf{Q}(u)$ in the imaginary time domain. Therefore, the cosine transformation described by Kresse and co-workers[15] is equivalent to the double-Laplace transformation in its integral formulation (eq 31), and we can follow the procedure of ref 15 to obtain optimized weights and roots for the transformation.

Additionally, the minimax grid for the imaginary frequency domain will be utilized for the final frequency integration replacing the Clenshaw−Curtis quadrature[32] to obtain the RPA correlation energy according to eq 28, which decreases the number of necessary quadrature points from 60−100 to 10−20.

2.3. Memory Efficient Calculation of the $\mathbf{F}_{\mathrm{INT}}$ Matrix. One of the major bottlenecks of the ω-CDD method is the huge memory requirement of the complete 3-center integrals in the AO basis. To find a strategy to reduce this memory requirement, we consider the formation of the $\mathbf{F}_{\mathrm{INT}}$ matrix according to[23]

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \mathrm{Tr}(\mathbf{Z}_M^{(p)T}\overline{\mathbf{P}}^{(p)}\mathbf{Z}_N^{(p)}) \tag{42}$$

where the pivoted Cholesky factorization of a given matrix $\mathbf{A}$ is abbreviated by $\mathbf{A} = \mathbf{L}\mathbf{L}^T$ and[23]

$$\mathbf{Z}_M^{(p)} = \mathbf{B}^M\underline{\mathbf{L}}^{(p)} \text{ with } \underline{\mathbf{P}}^{(p)} = \underline{\mathbf{L}}^{(p)}\underline{\mathbf{L}}^{(p)T} \tag{43}$$

The Cholesky factor $\underline{\mathbf{L}}^{(p)}$ is a transformation matrix to a local Cholesky basis,[30,31] and therefore, transformation of the 3-center integrals $\mathbf{B}^M$ to this local basis reduces the dimensions of each $\mathbf{B}^M$ from $N_{\mathrm{basis}} \times N_{\mathrm{basis}}$ to $N_{\mathrm{basis}} \times N_{\mathrm{occ}}$ while preserving all sparsity of the original matrix.[23] However, because $\mathbf{Z}_M^{(p)}$ is required for each Laplace point p, precontraction of each $\mathbf{B}^M$ with the Cholesky factors requires $N_{\mathrm{aux}}N_{\mathrm{basis}}N_{\mathrm{occ}}\tau$ memory, where τ is the number of Laplace points and, therefore, quickly becomes unfeasible.

Reformulating the calculation of $\mathbf{F}_{\mathrm{INT}}$ using the idempotency relation of the ground state density $\mathbf{P}$

$$\mathbf{P} = \mathbf{PSP} \tag{44}$$

with the 2-center overlap $\mathbf{S}$ and the extension to pseudodensity matrices

$$\underline{\mathbf{P}} = \underline{\mathbf{P}}\mathbf{S}\underline{\mathbf{P}} \tag{45}$$

according to

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \mathrm{Tr}(\underline{\mathbf{P}}^{(p)}\mathbf{B}^M\overline{\mathbf{P}}^{(p)}\mathbf{B}^N) \tag{46}$$

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \mathrm{Tr}(\underline{\mathbf{P}}^{(p)}\mathbf{SPB}^M\overline{\mathbf{P}}^{(p)}\mathbf{B}^N\mathbf{PS}) \tag{47}$$

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \mathrm{Tr}(\underline{\mathbf{P}}^{(p)}\mathbf{SL}\underline{\mathbf{L}}^T\mathbf{B}^M\overline{\mathbf{P}}^{(p)}\mathbf{B}^N\mathbf{L}\underline{\mathbf{L}}^T\mathbf{S}) \tag{48}$$

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \mathrm{Tr}(\mathbf{L}^T\mathbf{S}\underline{\mathbf{P}}^{(p)}\mathbf{SL}\underline{\mathbf{L}}^T\mathbf{B}^M\overline{\mathbf{P}}^{(p)}\mathbf{B}^N\mathbf{L}) \tag{49}$$

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \tilde{\underline{P}}_{\underline{j}\,\underline{i}}^{(p)}\tilde{B}_{\underline{i}\nu}^M\overline{P}_{\nu\mu}^{(p)}\tilde{B}_{\mu\underline{j}}^N \tag{50}$$

allows to precontract each $\mathbf{B}^M$ with the Cholesky factor $\mathbf{L}$ of the occupied one-particle density $\mathbf{P}$, which is not dependent on the Laplace points. Again, the dimensions are reduced from $N_{\mathrm{basis}} \times N_{\mathrm{basis}}$ to $N_{\mathrm{basis}} \times N_{\mathrm{occ}}$ while all sparsity of the original matrix is preserved, which will be shown in section 4. In this way, the required memory for saving the 3-center integrals is reduced by a factor of $\frac{N_{\mathrm{basis}}}{N_{\mathrm{occ}}}$ (with the total memory requirement being $N_{\mathrm{aux}}N_{\mathrm{basis}}N_{\mathrm{occ}}$), which is highly beneficial, especially for large basis sets. The final step for obtaining $\mathbf{F}_{\mathrm{INT}}$ is given by

$$(\mathbf{F}_{\mathrm{INT}}^{(p)})_{MN} = \tilde{\underline{B}}_{\underline{j}\mu}^{(p),M}\tilde{B}_{\mu\underline{j}}^N \tag{51}$$

As the reformulation of $\mathbf{F}_{\mathrm{INT}}$ presented in this section makes use of only the Cholesky decomposed ground state density (CDGD) $\mathbf{P}$, we will term the algorithm ω-CDGD-RPA.

2.4. Parallel Implementation. In the previous section, we showed that the memory requirements of the 3-center integral matrix can be decreased significantly by precontraction with the Cholesky factor of the ground state density. However, for large systems the 3-center integral matrix easily exceeds the available memory on a single node even with reduced dimensions. Therefore, our parallel implementation not only focuses on reducing calculation times of RPA correlation energies but also on a scalable decrease in memory requirements of the large 3-center integral matrix per node. The implementation presented here is a MPI/OpenMP hybrid parallelization of the ω-CDGD method, where OpenMP is mainly used for linear algebra and is therefore not discussed. The pseudocode of our implementation is shown in Figure 2.

The two most time-consuming steps in the ω-CDGD algorithm are, first, the formation of $\mathbf{F}_{\mathrm{INT}}$ according to eq 51 and, second, evaluation of the 3-center integrals $B_{\nu\sigma}^M$. For tackling all three of the bottlenecks mentioned above at once, a parallelization with respect to auxiliary basis functions is obvious. Therefore, each node calculates the 3-center integrals of a specific batch of RI basis functions (AUX_1...AUX_n) and

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

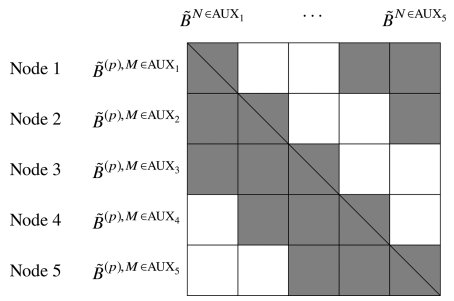


Figure 1. Schematic description of calculated unique blocks per node. Only the gray boxes are calculated; the white boxes are not calculated due to symmetry reasons.

keeps this part in memory (if enough memory is available) throughout the complete calculation. On each node, the 3-center integrals in the atomic orbital space are additionally evaluated in batches of RI basis functions, which are transformed into the Cholesky space by contraction with the Cholesky factor of the ground state density right after evaluation to reduce memory requirements. For evaluating the $\mathbf{F}_{\mathrm{INT}}$ matrix, the nodes are set up in a cyclic topology. Each node calculates its specific diagonal block of the matrix $((\mathbf{F}_{\mathrm{INT}})_{M\in\mathrm{AUX}_i,N\in\mathrm{AUX}_i})$ according to eq 51 and then sends a copy of its batch of 3-center integrals $(\tilde{B}_{\mu\underline{i}}^{N\in\mathrm{AUX}_i})$ to the neighboring node. In the next cycle, each node calculates the next unique part of the $\mathbf{F}_{\mathrm{INT}}$ matrix $((\mathbf{F}_{\mathrm{INT}})_{M\in\mathrm{AUX}_i,N\in\mathrm{AUX}_2})$ and passes the batch (AUX_2) on to the next neighbor. As $\mathbf{F}_{\mathrm{INT}}$ is symmetric, there are $\frac{n(n-1)}{2}$ unique nondiagonal parts of the matrix, where n is the total number of computing nodes, which is also the minimum number of send operations (see Figure 1). Note, however, that the send operations are conducted in parallel and the number of cycles $\left(N_{\mathrm{cycle}} = \frac{n-1}{2}\right)$ is of more interest.

To prevent repetitive recalculation of $\tilde{\underline{B}}_{\underline{j}\mu}^{(p),M}$ in eq 51 or increasing the number of cycles by a factor of τ, a variable number of these matrices are precomputed. For computational efficiency, all τ matrices should be precomputed. However, to reduce memory requirements, it would be best to precompute only one of these matrices. In our implementation, the complete available memory is exploited to precompute as many of these matrices as possible. In the case of dense matrices, the approximation of the required memory is trivial because the memory requirements of all matrices are known beforehand. To approximate the required memory in case of sparse matrices, we use the fact that $\tilde{\underline{B}}_{\underline{j}\mu}^{(p),M}$ becomes less sparse when the value of the respective Laplace point p becomes

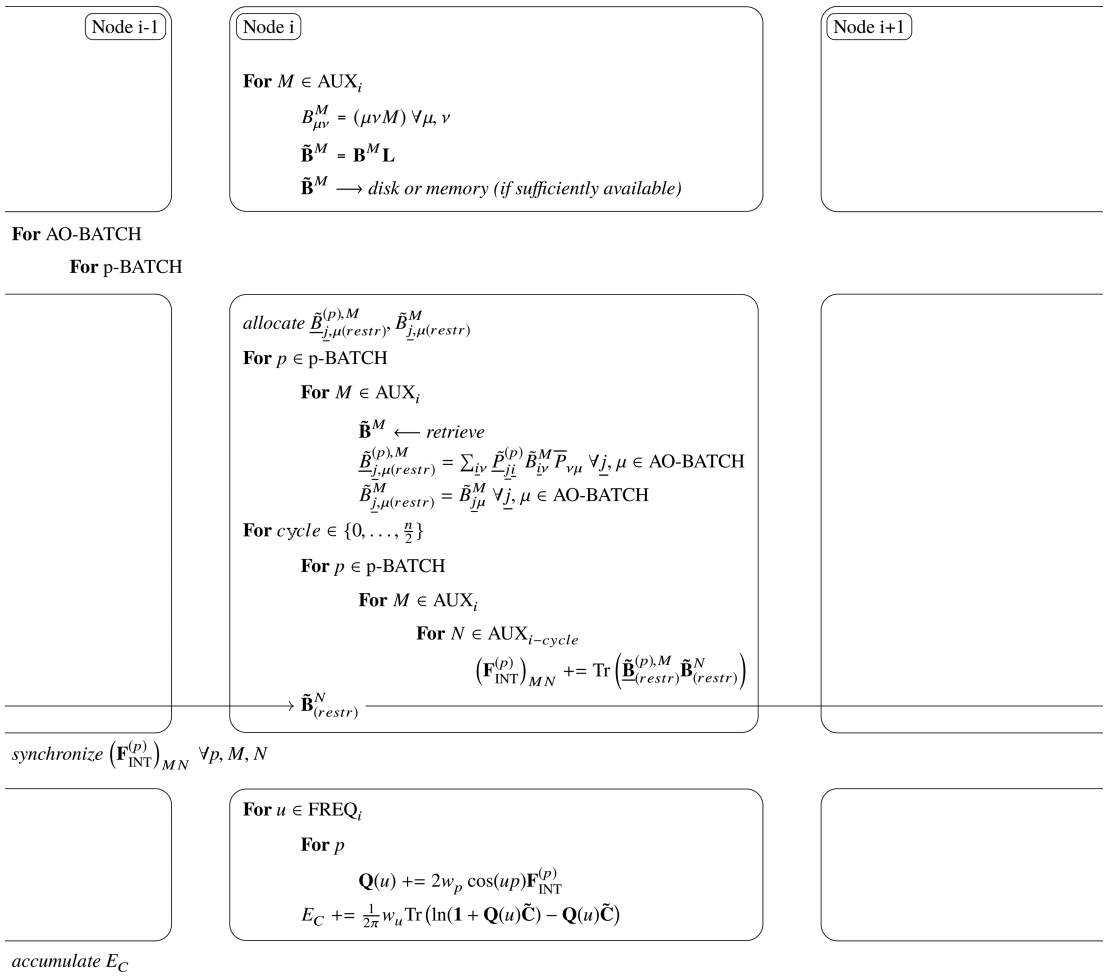


Figure 2. Pseudocode for the parallel calculation of the RPA correlation energy.

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

Table 1. Mean Absolute Deviations (MAD) and Maximum Absolute Deviations (MAX) of the Calculated Absolute Energies (Upper Part) as well as the Interaction Energies (Lower Part) Obtained with the ω-CDD and the New ω-CDGD Method from the Reference MO-RI-RPA Calculations for the S66 (Left) and L7 (Right) Test Sets

		S66		L7	
		MAD [H]	MAX [H]	MAD [H]	MAX [H]
absolute energies	ω-CDD	2×10^{-03}	1×10^{-02}	9×10^{-03}	4×10^{-02}
	ω-CDGD	7×10^{-07}	5×10^{-06}	9×10^{-07}	1×10^{-05}
interaction energies	ω-CDD	5×10^{-04}	3×10^{-03}	3×10^{-03}	1×10^{-02}
	ω-CDGD	9×10^{-07}	5×10^{-06}	2×10^{-06}	1×10^{-05}

smaller. Therefore, we calculate $\underline{\tilde{B}}^{(p),M}_{j\mu}$ with the smallest value of p and approximate the required memory by considering the number of allocated blocks of all necessary matrices. Note that we do not need to recalculate this matrix in the following evaluation of $\mathbf{F}_{\mathrm{INT}}$ and, hence, this approximation does not decrease the efficiency of our algorithm. The final frequency integration (eq 28) is parallelized with respect to frequency points, which necessitates having the complete $\mathbf{F}_{\mathrm{INT}}$ matrix on all nodes because $\mathbf{Q}(u)$ has to be computed on the fly from all $\mathbf{F}^{(p)}_{\mathrm{INT}}$ according to eq 31. For the complete matrix to be formed on all nodes, the rotary scheme described above is again used. This time, however, at least $n - 1$ cycles are necessary.

The parallel implementation described so far reduces the memory requirements of the large 3-center matrix by a factor of $\frac{n}{2}$ because two batches of $\tilde{B}^{N}_{\mu\underline{i}}$ are necessary on every node. To further reduce the required memory of the algorithm (if necessary), we implemented an additional loop over batches of atomic orbitals (AO_1...AO_c) on each node. Note that in this case batching with respect to atomic orbitals is superior because each batch ($\tilde{B}^{(p),M\in\mathrm{AUX}_1}_{\underline{i}\ \mu\in\mathrm{AO}_1}$) only needs to be combined with the same AO batch of matrices ($\tilde{B}^{N\in\mathrm{AUX}_n}_{\mu\in\mathrm{AO}_1,\underline{i}}$) and not with all other AO batches because only the trace over atomic orbitals is necessary in the calculation of $\mathbf{F}_{\mathrm{INT}}$. The final result for the unique part of $\mathbf{F}_{\mathrm{INT}}$ ($(\mathbf{F}_{\mathrm{INT}})_{M\in\mathrm{AUX}_1,N\in\mathrm{AUX}_n}$) is then obtained by summing all results of the c AO batches.

3. COMPUTATIONAL DETAILS

The new ω-CDGD method as well as the ω-CDD[23] and the MO-RI-RPA[13] methods were implemented in the FermiONs++ program package.[41,42] Kohn–Sham orbitals used for the RPA energy calculations were obtained by preceding DFT calculations employing the generalized gradient approximation of Perdew, Burke, and Ernzerhof[43,44] (PBE) with def2-SVP, def2-TZVP, and def2-QZVP basis sets.[45,46] The RI approximation, which is only used for 4-center integrals in the correlation part of the RPA energy, uses the corresponding auxiliary basis sets[47,48] with the attenuated Coulomb metric with attenuation parameter $\omega = 0.1$. For the Laplace expansion in the AO implementations, 13–15 quadrature points were used. In the case of the ω-CDD algorithm, pretabulated values for the weights and roots of ref 49 are employed. Within the ω-CDGD method, we implemented the *sloppy Remez*[15] algorithm to obtain optimized weights for the integrated double-Laplace expansion to switch between the representation of $\mathbf{Q}$ in the imaginary time domain to the representation in the imaginary frequency domain. The MO-RI-RPA as well as the ω-CDD methods use the Clenshaw–Curtis scheme[13] with 60–100 quadrature points for the final frequency integration, whereas the ω-CDGD algorithm utilizes a minimax grid[15] with 13–15 frequency points.

For obtaining the minimax grids for the imaginary time and frequency domain, a Remez algorithm as described by Kresse and co-workers[15] was implemented. As a starting guess for the imaginary time domain, pretabulated values by Hackbusch and co-workers[49] are used. For obtaining a starting guess for the weights and roots in the imaginary frequency domain, a least-squares fit was performed utilizing the Levenberg–Marquardt algorithm[50] after the starting values were initialized randomly.

Total energies were obtained by adding the correlation energy to the exact Hamiltonian expectation value calculated from the PBE reference orbitals. Core orbitals were frozen in all RPA calculations.

For accuracy benchmarks, the full S66[51] test set of small-molecule interaction energies and the L7[52] test set of dispersion-dominated molecules of larger size were used. Investigations on performance and scaling behavior were conducted with a test set of linear n-alkanes and DNA fragments of increasing size. For the efficiency of the parallel implementation to be tested, the L7 test set was used again because the contained molecules can be seen as a representative selection of molecular sizes (15–112 atoms) for many applications. As an illustrative example for the applicability of our new method, the layer distance between two covalent organic framework (COF) pores was calculated.

4. RESULTS AND DISCUSSION

4.1. Accuracy: S66 and L7 Test Sets. First, the accuracy of the newly implemented ω-CDGD method is investigated and compared to that of the ω-CDD method.[23] Therefore, the full S66[51] test set of small-molecule interaction energies and the L7[52] benchmark set of dispersion-dominated molecules of larger size were calculated.

As reference serves our implementation of the canonical MO-RI-RPA method described by Furche and co-workers[13] using the Clenshaw–Curtis quadrature with optimization of the scaling parameter and 60 quadrature points. For the Laplace expansion in the two AO implementations, 15 quadrature points were employed. Correspondingly, 15 quadrature points were used for the final frequency integration in the ω-CDGD-RI-RPA method, and the ω-CDD algorithm used the Clenshaw–Curtis quadrature with settings equal to those of the MO-RI-RPA calculations. Note that the ω-CDD method yields better results for interaction energies when a fixed integration interval of $u \in [0;400]$ au is used.[23] However, for comparison reasons, the same settings for the frequency integration were used as for the reference calculations. The calculations of the full S66 test set were performed using the def2-QZVP basis set, whereas for the L7 test set, the def2-TZVP basis set was employed. All other settings remained equal in all calculations to facilitate comparison.

Table 1 shows the mean absolute deviation (MAD) and the maximum absolute deviation (MAX) of the calculated

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505–2515

correlated energies (upper part) as well as the interaction energies (lower part) using the two AO implementations from the MO reference calculations for both test sets. Starting with the S66 test set, the ω-CDGD method shows a significantly increased accuracy in absolute energies by 4 orders of magnitude with respect to both deviation measures due to the optimized integration schemes. Considering the interaction energies, the gain in accuracy is less dramatic but still lies within 3 orders of magnitude for the mean absolute as well as the maximum absolute deviation. Similar observations can be made when considering the L7 test set. The observed accuracy in the calculation of absolute energies is again increased significantly using the new ω-CDGD method. For both measures of deviation from the reference results, the occurring error is decreased by at least 3 orders of magnitude. As for the S66 test set, the difference in accuracy is smaller between the ω-CDD and the ω-CDGD implementation when referring to interaction energies. Still, the mean absolute as well as the maximum absolute deviation from the reference are lowered by 3 orders of magnitude.

4.2. Performance and Scaling Behavior. Next, the performance improvements and the scaling behavior of our new method are investigated in comparison with the preceding ω-CDD-RPA method. To do so, we calculated RPA correlation energies of a set of linear *n*-alkanes and DNA fragments of increasing size. The calculations on linear *n*-alkanes using the ω-CDD algorithm were conducted with 13 quadrature points for the Laplace expansion and 100 quadrature nodes for the frequency integration on a fixed integration interval of $u \in [0;300]$ au. The ω-CDGD method used 13 points for the frequency integration as well as the integrated double-Laplace expansion. For the calculations on the DNA fragments, we employed 15 instead of 13 quadrature points for the Laplace expansion as well as the minimax grid-based frequency integration. All calculations employed the def2-SVP basis set and were performed on an Intel Xeon E5-2667 processor using 16 threads.

4.2.1. Linear n-Alkanes. As a first test, we calculated a set of linear *n*-alkanes of increasing length. The results are shown in Table 2. Note that the calculations up to $C_{80}H_{162}$ were performed with standard dense matrix algebra, whereas all calculations of alkanes with larger size were conducted using sparse matrix algebra.

Considering the timings shown in Table 2, it becomes obvious that the newly implemented ω-CDGD method is asymptotically 4-times faster than the ω-CDD method for dense as well as sparse matrix algebra. This speed-up can be explained by the fact that, in the case of the ω-CDGD method, only the $\mathbf{F}_{INT}$ matrix needs to be calculated, whereas in the ω-CDD algorithm, all three expensive $\mathbf{F}$ matrices ($\mathbf{F}_0$, $\mathbf{F}_{INT}$, and $\mathbf{F}_D$) need to be evaluated. As the formation of $\mathbf{F}_0$ is independent of the Laplace points *p*, the time consumption is almost negligible compared to those of the other two $\mathbf{F}$ matrices. However, because the formation of $\mathbf{F}_D$ is ~3-times as expensive as the formation of $\mathbf{F}_{INT}$ (see ref 23), a speed-up of approximately 4 is to be expected when only $\mathbf{F}_{INT}$ is computed.

A linear plot of the wall times against the number of AO basis functions is shown in Figure 3 (left). As can be seen, both implementations show a linear increase in the wall times for calculations using sparse matrix algebra (more than 1930 basis functions). In addition, the dashed line in the graph shows the wall times for the cubic-scaling frequency integration in the ω-CDGD algorithm. The contribution to the total wall time is very small even for the largest system under investigation (902 atoms) due to the small prefactor. On the right-hand side of Figure 3, a log−log plot of the wall times against the number of basis functions is shown. It shows that the scaling behavior for small as well as large systems remains roughly the same in the new implementation, however, with a significant decrease in the prefactor compared to the original implementation. As already stated above, an effective linear-scaling behavior can be observed for systems having more than 1930 basis functions in both algorithms.

Table 2. Wall Times for the Calculation of RI-RPA Correlation Energies of Linear *n*-Alkanes with Increasing Length Using the ω-CDD and ω-CDGD Methods with the Last Column Giving the Speed-up of the Computation Time Using the ω-CDGD Method Compared to the ω-CDD Method

			time [s]		
	molecule	N_{basis}	ω-CDD	ω-CDGD	speed-up
dense	C_5H_{12}	130	1	1	1.0
	$C_{10}H_{22}$	250	5	2	2.5
	$C_{15}H_{32}$	370	17	5	3.4
	$C_{20}H_{42}$	490	43	12	3.6
	$C_{30}H_{62}$	730	168	42	4.0
	$C_{40}H_{82}$	970	459	111	4.1
	$C_{50}H_{102}$	1210	978	232	4.2
	$C_{60}H_{122}$	1450	1887	443	4.3
	$C_{70}H_{142}$	1690	3221	763	4.2
	$C_{80}H_{162}$	1930	5329	1241	4.3
sparse	$C_{90}H_{182}$	2170	7111	1860	3.8
	$C_{100}H_{202}$	2410	7914	2072	3.8
	$C_{120}H_{242}$	2890	11348	3008	3.8
	$C_{160}H_{322}$	3850	15351	3974	3.9
	$C_{200}H_{402}$	4810	21444	5564	3.9
	$C_{300}H_{602}$	7210	40239	10003	4.0

4.2.2. DNA Fragments. To further investigate the performance improvements of our new method, we calculated RPA correlation energies of DNA fragments of increasing size. All calculations were performed using dense matrix algebra because no performance improvements were observed for the fairly moderate system sizes when sparse matrix algebra was applied. The results are shown in Table 3 and visualized in Figure 4. As for the set of linear *n*-alkanes, the ω-CDGD method is asymptotically 4-times faster than the ω-CDD algorithm. Note that the wall time for the ω-CDD-RPA calculation on the four base pair DNA fragment (DNA_4) was extrapolated because the memory requirements of the large 3-center integral matrix exceeded the available memory.

In conclusion, the ω-CDGD algorithm, which only makes use of the $\mathbf{F}_{INT}$ matrix calculated by just decomposing the ground state density, preserves sparsity as well as the ω-CDD method while at the same time reducing the memory requirements by a factor of $\frac{N_{basis}}{N_{occ}}$ and the run-time prefactor by a factor of 4.

4.3. Parallel Implementation. As described in section 2.4, a parallelization of the ω-CDGD method was implemented to further reduce the memory requirement of the algorithm and to speed up the calculation of RPA correlation energies to open the way for investigations of larger systems. In the following, the parallel efficiency of the implementation, defined as the observed speed-up divided by the number of nodes, is

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

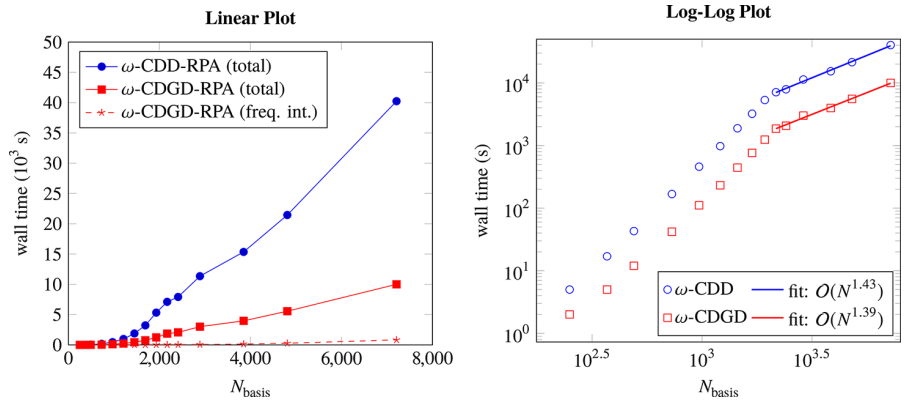


Figure 3. Linear plot (left) and the respective log−log plot (right) of the wall times for calculating RI-RPA correlation energies of linear *n*-alkanes using the ω-CDD (blue) and ω-CDGD (red) methods against the number of basis functions. Additionally, the wall time for the final frequency integration in the ω-CDGD method is shown (red, dashed). The log−log plot further shows linear fits for the ω-CDD (blue line) and ω-CDGD (red line) methods. For the linear fits, only data points are used, which were calculated using sparse matrix algebra.

Table 3. Wall Times for the Calculation of RI-RPA Correlation Energies of DNA Fragments with Increasing Size Using the ω-CDD and ω-CDGD Methods with the Last Column Giving the Speed-up of the Computation Time Using the ω-CDGD Method Compared to the ω-CDD Method[a]

		time [s]		
molecule	N_{basis}	ω-CDD	ω-CDGD	speed-up
DNA_1	625	126	35	3.6
DNA_2	1332	2018	528	3.8
DNA_4	2746	*32288	7755	4.2

[a]The value marked with an asterisk (*) was extrapolated conservatively.

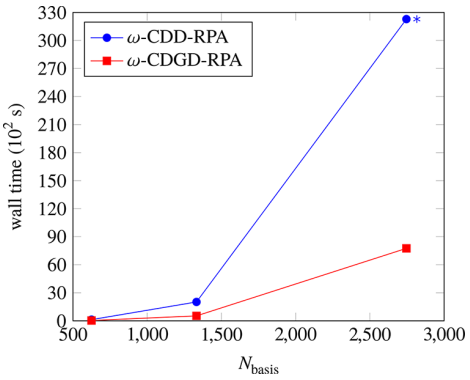


Figure 4. Linear plot of the wall times for calculating RI-RPA correlation energies of DNA fragments (DNA_1, DNA_2, and DNA_4) using the ω-CDD (blue) and ω-CDGD (red) methods against the number of basis functions. The data point marked with an asterisk (*) was extrapolated conservatively.

investigated. To examine the changes in the parallel efficiency by varying the number of computation nodes, correlated energies of the complete L7 test set[52] were calculated on 1, 5, and 10 nodes using the def2-TZVP basis set on dual-core processor Intel Xeon E5-2620 machines using 12 threads per node. Note that these calculations do not only include the RPA correlation energy but also the Hamiltonian expectation value to give the total correlated energy of the system. The results are shown in Table 4.

Table 4. Cumulative Wall Times, Speed-ups, and Parallel Efficiencies for the Calculation of Correlated Energies of the Complete L7 Test Set Using 1, 5, and 10 Nodes

number of nodes	time [s]	speed-up	efficiency
1	33343		
5	6824	4.9	0.98
10	3840	8.7	0.87

As can be seen, using 5 compute nodes results in a speed-up in the cumulative calculation time of the complete L7 test set by 4.9, which corresponds to a parallel efficiency of 98%. On 10 computing nodes, however, the parallel efficiency decreases to 87%. This can be explained by the increase in communication over the network, which is particularly a problem for smaller systems where the ratio between communication and calculation time is large. However, because the calculation times for the relatively small systems in the L7 test set are very short even on only 1 node, the lower parallel efficiency for 10 nodes is less significant.

Focusing on larger systems, e.g., the circumcoronene guanine-cytosine base pair complex (Table 5), which is the system with the largest number of basis functions in the test set, shows that even with 10 nodes a parallel efficiency of 96% is observed. This means that, especially for large systems for which the parallel efficiency matters most, significant speed-ups are observed.

Table 5. Wall Times, Speed-ups, and Parallel Efficiencies for Calculating the Correlated Energy of the Circumcoronene Guanine-Cytosine Base Pair Complex (101 Atoms, 2431 Basis Functions, 5968 Auxiliary Basis Functions) from the L7 Test Set Using 1−10 Computing Nodes

number of nodes	time [s]	speed-up	efficiency
1	10074		
2	4974	2.0	1.00
4	2582	3.9	0.98
6	1673	6.0	1.00
8	1280	7.9	0.99
10	1046	9.6	0.96

4.4. Illustrative Application. Efficient carbon capture and storage as well as atmospheric water capture are important

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

societal challenges and necessitate the development of materials with specific sorption properties. In 2015, Stegbauer et al.[53] reported that COFs show very high CO_2 and water uptake capacities at low pressures, making them highly interesting for environmental applications.

For understanding the absorption characteristics of COFs, detailed structural analyses are indispensable. One of the properties of intereset is the distance between two layers of the COF. As the interlayer distance is strongly influenced by dispersion interactions, a good description of these effects is essential for obtaining reliable computational results. We calculated total RPA energies of two azine-benzene-COF (AB-COF, see Figure 5) pores with eclipsed stacking (384 atoms) and distances between the two pores ranging from 3.2 to 4.0 Å using our ω-CDGD algorithm. All calculations were conducted using dense matrix algebra because no useful sparsity could be expected due to the highly delocalized electronic structure of the system. Note that the use of dense linear algebra prevents linear-scaling behavior, which relies on the sparsity of density matrices. The results are shown in Figure 6.

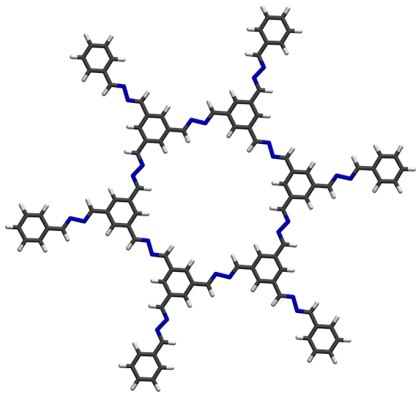


Figure 5. Structure representation of one AB-COF pore.

Although the preceding PBE calculations do not show a minimum in the calculated range with both the def2-SVP as well as the def2-TZVP basis sets, the RPA calculations clearly improve upon the PBE results and show a minimum at ∼3.5 Å using the def2-SVP basis set and a slightly right-shifted minimum at 3.6 Å with the def2-TZVP basis set. Both results for this model system are in good agreement with the experimental value of Stegbauer et al.[53] who reported an interlayer distance of 3.44 Å. As expected, the DFT calculations using the PBE functional are not sufficient to obtain reliable results for the interlayer distance of the COF due to the missing description of dispersion effects. However, with the methodology presented in this work, we are now able to correctly describe noncovalent interactions without necessitating additional parameters.

5. CONCLUSION

The memory bottleneck of our first linear-scaling RPA formulations has been overcome by a reformulation of the $\mathbf{F}_{\mathrm{INT}}$ matrix, which only uses the Cholesky factors of the ground state density. This new formulation preserves sparsity as well as the previous Cholesky decomposition of the pseudodensities method but reduces the memory requirements of the large 3-center integral matrix by a factor of $\frac{N_{\mathrm{basis}}}{N_{\mathrm{occ}}}$. This is essential to further extend the applicability of this method to ever larger systems. Additionally, we have presented a parallel implementation of our method, which is shown to be highly efficient and also enables a further scalable decrease in memory requirements, opening the door for investigations of large molecules even on small- to medium-sized computing clusters. While our previous formulations required two different equations for the calculation of $\mathbf{Q}(u)$, which necessitates evaluating three different forms of $\mathbf{F}$ matrices, we have overcome this issue by utilizing minimax grids for the imaginary time and frequency domain as well as an optimized transform scheme to switch between the two domains. To this end, we have shown that $\mathbf{F}_{\mathrm{INT}}$ is the representation of $\mathbf{Q}(u)$ in the imaginary time domain and that the integrated double-Laplace expansion is equivalent to a Fourier transform of the frequency-dependent noninteracting polarization propagator into the imaginary time domain. Employing optimized weights and roots for this transformation thus allows for using only the "INT"-formulation (eq 31). This decreases the run-time prefactor of our algorithm by a factor of 4 and at the same time increases the accuracy of our method by up to 4 orders of magnitude. Replacing the Clenshaw−Curtis scheme for the final frequency

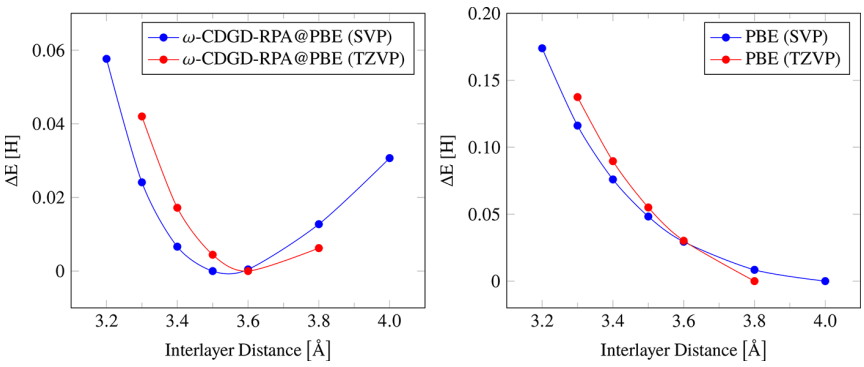


Figure 6. Plot of the relative energies of the AB-COF pore dimer calculated with the ω-CDGD method using PBE reference orbitals (left) and the PBE functional (right) employing the def2-SVP (blue) as well as the def2-TZVP (red) basis set against the interlayer distance. All data points were referenced with respect to the lowest computed value. The total wall time for each RPA correlation energy calculation using the def2-TZVP basis set is 130014 s on 10 computing nodes. The time for the communication between the nodes is 8307 s, which is approximately 6% of the total calculation time.

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

integration with a minimax quadrature decreases the number of quadrature points from 60−100 to 10−20 without losing accuracy. In conclusion, we have described an accurate linear-scaling RPA theory that is efficiently parallelized, has a low prefactor, is numerically very stable, and has low memory requirements that can be further lowered by distribution over multiple compute nodes. This enables calculations of large molecular systems in a fraction of the time of former theories on compute clusters typically available in local research groups. As an illustrative application, we used our new method to calculate the distance between two COF pores comprising 384 atoms and found the equilibrium distance to be in good agreement with experimental results.

AUTHOR INFORMATION

Corresponding Author

*E-mail: christian.ochsenfeld@uni-muenchen.de.

ORCID

Gökcen Savasci: 0000-0002-6183-7715

Christian Ochsenfeld: 0000-0002-4189-6558

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support was provided by the Excellence Cluster EXC114 (CIPSM) and the SFB749 by the Deutsche Forschungsgemeinschaft (DFG). C.O. acknowledges, in addition, financial support as a Max-Planck Fellow at the MPI-FKF Stuttgart.

REFERENCES

(1) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315−2372.

(2) Kristyán, S.; Pulay, P. Can (semi)local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.* **1994**, *229*, 175−180.

(3) Pérez-Jordá, J.; Becke, A. A density-functional study of van der Waals forces: rare gas diatomics. *Chem. Phys. Lett.* **1995**, *233*, 134−137.

(4) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.

(5) Bohm, D.; Pines, D. A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas. *Phys. Rev.* **1953**, *92*, 609−625.

(6) Dobson, J. *Time-Dependent Density Functional Theory*; Springer: Berlin, Germany, 2006; Vol. *706*.

(7) Fuchs, M.; Niquet, Y. M.; Gonze, X.; Burke, K. Describing static correlation in bond dissociation by Kohn-Sham density functional theory. *J. Chem. Phys.* **2005**, *122*, 094116.

(8) Harl, J.; Kresse, G. Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 1−8.

(9) Kresse, G.; Harl, J. Accurate Bulk Properties from Approximate Many-Body Techniques. *Phys. Rev. Lett.* **2009**, *103*, 4−7.

(10) Langreth, D. C.; Perdew, J. P. The Exchange-Correlation Energy of a Metallic Surface. *Solid State Commun.* **1975**, *17*, 1425−1429.

(11) Lume, V. O.; Langreth, D. C.; Perdew, J. P. Exchange-correlation energy of a metallic surface: Wave-vector analysis*. *Phys. Rev. B* **1977**, *15*, 2884−2901.

(12) Furche, F. Developing the random phase approximation into a practical post-Kohn-Sham correlation model. *J. Chem. Phys.* **2008**, *129*, 114105.

(13) Eshuis, H.; Yarkony, J.; Furche, F. Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration. *J. Chem. Phys.* **2010**, *132*, 234114.

(14) Eshuis, H.; Furche, F. Basis set convergence of molecular correlation energy differences within the random phase approximation. *J. Chem. Phys.* **2012**, *136*, 084105.

(15) Kaltak, M.; Klimeš, J.; Kresse, G. Low Scaling Algorithms for the Random Phase Approximation: Imaginary Time and Laplace Transformations. *J. Chem. Theory Comput.* **2014**, *10*, 2498−2507.

(16) Kállay, M. Linear-scaling implementation of the direct random-phase approximation. *J. Chem. Phys.* **2015**, *142*, 204105.

(17) Schurkus, H. F.; Ochsenfeld, C. Communication: An effective linear-scaling atomic-orbital reformulation of the random-phase approximation using a contracted double-Laplace transformation. *J. Chem. Phys.* **2016**, *144*, 031101.

(18) Pulay, P. Localizability of Dynamic Electron Correlation. *Chem. Phys. Lett.* **1983**, *100*, 151−154.

(19) Kapuy, E.; Csépes, Z.; Kozmutza, C. Application of the Many-Body Perturbation Theory by Using Localized Orbitals. *Int. J. Quantum Chem.* **1983**, *23*, 981−990.

(20) Förner, W.; Ladik, J.; Otto, P.; Cížek, J. Coupled-Cluster Studies. II. The Role of Localization in Correlation Calculations on Extended Systems. *Chem. Phys.* **1985**, *97*, 251−262.

(21) Wilhelm, J.; Seewald, P.; Del Ben, M.; Hutter, J. Large-Scale Cubic-Scaling Random Phase Approximation Correlation Energy Calculations Using a Gaussian Basis. *J. Chem. Theory Comput.* **2016**, *12*, 5851−5859.

(22) Vahtras, O.; Almlöf, J.; Feyereisen, M. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514−518.

(23) Luenser, A.; Schurkus, H. F.; Ochsenfeld, C. Vanishing-Overhead Linear-Scaling Random Phase Approximation by Cholesky Decomposition and an Attenuated Coulomb-Metric. *J. Chem. Theory Comput.* **2017**, *13*, 1647−1655.

(24) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. Auxiliary basis expansions for large-scale electronic structure calculations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6692−6697.

(25) Jung, Y.; Shao, Y.; Head-Gordon, M. Fast Evaluation of Scaled Opposite Spin Second-Order Møller-Plesset Correlation Energies Using Auxiliary Basis Expansions and Exploiting Sparsity. *J. Comput. Chem.* **2007**, *28*, 1953−1964.

(26) Reine, S.; Tellgren, E.; Krapp, A.; Kjærgaard, T.; Helgaker, T.; Jansik, B.; Høst, S.; Salek, P. Variational and robust density fitting of four-center two-electron integrals in local metrics. *J. Chem. Phys.* **2008**, *129*, 104101.

(27) Koch, H.; Sánchez De Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481−9484.

(28) Higham, N. J. Cholesky factorization. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 251−254.

(29) Harbrecht, H.; Peters, M.; Schneider, R. On the low-rank approximation by the pivoted Cholesky decomposition. *Appl. Numer. Math.* **2012**, *62*, 428−440.

(30) Zienau, J.; Clin, L.; Doser, B.; Ochsenfeld, C. Cholesky-decomposed densities in Laplace-based second-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2009**, *130*, 204112.

(31) Maurer, S. A.; Clin, L.; Ochsenfeld, C. Cholesky-decomposed density MP2 with density fitting: Accurate MP2 and double-hybrid DFT energies for large systems. *J. Chem. Phys.* **2014**, *140*, 224112.

(32) Boyd, J. P. Exponentially Convergent Fourier-Chebshev Quadrature Schemes on Bounded and Infinite Intervals. *J. Sci. Comput.* **1987**, *2*, 99−109.

(33) Einstein, A. Die Grundlage der allgemeinen Relativitätstheorie. *Ann. Phys.* **1916**, *354*, 769−822.

(34) Gunnarsson, O.; Lundqvist, B. I. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B* **1976**, *13*, 4274−4298.

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

(35) Niquet, Y. M.; Fuchs, M.; Gonze, X. Exchange-correlation potentials in the adiabatic connection fluctuation-dissipation framework. *Phys. Rev. A: At., Mol., Opt. Phys.* **2003**, *68*, 032507.
(36) Furche, F.; Van Voorhis, T. Fluctuation-dissipation theorem density-functional theory. *J. Chem. Phys.* **2005**, *122*, 164106.
(37) Ullrich, C.-A. *Time-Dependent Density-Functional Theory - Concepts and Applications*; Oxford University Press: Oxford, U.K., 2012.
(38) Grundei, M. M.; Burow, A. M. Random Phase Approximation for Periodic Systems Employing Direct Coulomb Lattice Summation. *J. Chem. Theory Comput.* **2017**, *13*, 1159−1175.
(39) Bates, J. E.; Furche, F. Communication: Random phase approximation renormalized many-body perturbation theory. *J. Chem. Phys.* **2013**, *139*, 171103.
(40) Mussard, B.; Rocca, D.; Jansen, G.; Ángyán, J. G. Dielectric Matrix Formulation of Correlation Energies in the Random Phase Approximation: Inclusion of Exchange Effects. *J. Chem. Theory Comput.* **2016**, *12*, 2191−2202.
(41) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, 134114.
(42) Kussmann, J.; Ochsenfeld, C. Preselective Screening for Linear-Scaling Exact Exchange-Gradient Calculations for Graphics Processing Units and General Strong-Scaling Massively Parallel Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 918−922.
(43) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396−1396.
(44) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(45) Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple zeta valence quality for atoms H-Kr. *J. Chem. Phys.* **2003**, *119*, 12753−12762.
(46) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
(47) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143−152.
(48) Hättig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple-ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59−66.
(49) Takatsuka, A.; Ten-No, S.; Hackbusch, W. Minimax approximation for the decomposition of energy denominators in Laplace-transformed Møller-Plesset perturbation theories. *J. Chem. Phys.* **2008**, *129*, 044112.
(50) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes 3rd ed.: The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: New York, NY, USA, 2007.
(51) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(52) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364−3374.
(53) Stegbauer, L.; Hahn, M. W.; Jentys, A.; Savasci, G.; Ochsenfeld, C.; Lercher, J. A.; Lotsch, B. V. Tunable Water and CO_2 Sorption Properties in Isostructural Azine-Based Covalent Organic Frameworks through Polarity Engineering. *Chem. Mater.* **2015**, *27*, 7874−7881.

DOI: 10.1021/acs.jctc.8b00177
J. Chem. Theory Comput. 2018, 14, 2505−2515

3.3 Publication III: Low-scaling analytical gradients for the direct random phase approximation using an atomic orbital formalism

M. Beuerle, C. Ochsenfeld,
"Low-Scaling Analytical Gradients for the Direct Random Phase Approximation using an Atomic Orbital Formalism",
J. Chem. Phys., **149**, 244111 (2018).

Abstract: We present an atomic orbital formalism to obtain analytical gradients within the random phase approximation for calculating first-order properties. Our approach allows exploiting sparsity in the electronic structure in order to reduce the computational complexity. Furthermore, we introduce Cholesky decomposed densities to remove the redundancies present in atomic orbital basis sets, making our method a competitive alternative to canonical theories also for small molecules. The approach is presented in a general framework that allows extending the methodology to other correlation methods. Beyond showing the validity and accuracy of our approach and the approximations used in this work, we demonstrate the efficiency of our method by computing nuclear gradients for systems with up to 600 atoms and 5000 basis functions.

THE JOURNAL OF CHEMICAL PHYSICS **149**, 244111 (2018)

Low-scaling analytical gradients for the direct random phase approximation using an atomic orbital formalism

Matthias Beuerle and Christian Ochsenfeld[a)]
Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 München, Germany and Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5–13, D-81377 München, Germany

(Received 20 August 2018; accepted 4 December 2018; published online 27 December 2018)

We present an atomic orbital formalism to obtain analytical gradients within the random phase approximation for calculating first-order properties. Our approach allows exploiting sparsity in the electronic structure in order to reduce the computational complexity. Furthermore, we introduce Cholesky decomposed densities to remove the redundancies present in atomic orbital basis sets, making our method a competitive alternative to canonical theories also for small molecules. The approach is presented in a general framework that allows extending the methodology to other correlation methods. Beyond showing the validity and accuracy of our approach and the approximations used in this work, we demonstrate the efficiency of our method by computing nuclear gradients for systems with up to 600 atoms and 5000 basis functions. *Published by AIP Publishing.*
https://doi.org/10.1063/1.5052572

I. INTRODUCTION

Density-functional-theory (DFT) is the most widely used method in quantum chemistry to calculate energetics and properties. Its success stems from the good compromise of computational cost and accuracy. There are, however, certain problems that come with the use of DFT. The most important are the limited transferability of highly parameterized functionals to a broad range of chemistry (see, e.g., Ref. 1) and the lacking description of dispersion effects.[2,3] For the latter problem, the most popular solution is the use of empirical dispersion corrections (see, e.g., Ref. 4). An alternative to DFT are wavefunction based methods such as second-order Møller-Plesset (MP2) perturbation theory[5] or coupled cluster (CC) methods,[6,7] which are significantly more computationally involved.

In recent years, the direct random phase approximation (RPA) (for an introduction see, e.g., Refs. 8–10) has regained popularity as a post-Kohn-Sham (KS) functional, due to its good description of dispersion[11,12] and the ability to describe small-gap systems, where other methods, e.g., MP2, fail drastically. Furthermore, the availability of algorithms that allow calculating RPA energies with low computational cost contributes to the popularity of the RPA. This was pioneered by Furche and co-workers,[13] who presented a resolution-of-the-identity (RI) RPA approach with $\mathcal{O}(N^4)$ computational cost, where N denotes the molecule size. Furthermore, linear-scaling methods to calculate RPA correlation energies were presented,[14–17] allowing for calculations on systems with more than 1000 atoms making it an attractive alternative to conventional DFT.

For the applicability of an electronic structure method, the access to properties beyond ground state energies is indispensable. Analytical gradients are the first step in this direction, where computational efficiency is key as well. To this end, $\mathcal{O}(N^6)$ schemes based on the CC variant of RPA have been presented,[18,19] as well as an $\mathcal{O}(N^4)$ RI-RPA gradient approach by Furche and co-workers.[20] Very recently Kresse and co-workers[21] presented an $\mathcal{O}(N^3)$ gradient approach for periodic systems using a plane wave basis set.

In this work, we extend our linear and low-scaling atomic orbital (AO) RI-RPA schemes[14–16] to first order properties taking one step further in the direction of making RPA a general purpose non-empirical alternative to conventional DFT.

II. THEORY

We formulate our theory starting from the AO-RI-RPA formalism,[14] which proved to be highly efficient for the calculation of RPA correlation energies. The Einstein sum convention is used throughout. Integration over spatial coordinates is avoided by representing all quantities in atomic and auxiliary basis sets and using discrete summations. We denote quantities related to occupied molecular orbitals (MOs) as $i, j, \ldots$ and those related to virtual orbitals $a, b, \ldots$. Atomic orbitals are enumerated as $\mu, \nu, \ldots$, Cholesky orbitals as $\underline{i}, \underline{j}, \ldots$, and auxiliary basis functions as $P, Q, \ldots$. Two-, three-, and four-center electron repulsion integrals (ERIs) will be denoted in the Mulliken notation.

A. Atomic orbital RPA total energies

In the adiabatic-connection formalism,[22] the total energy is usually calculated as

$$E = E_h[\{\rho_{\mathrm{KS}}\}] + E_J[\{\rho_{\mathrm{KS}}\}] + E_K[\{\rho_{\mathrm{KS}}\}] + E_c, \quad (1)$$

a)christian.ochsenfeld@uni-muenchen.de

where the sum of the one-particle energy contributions E_h and the Coulomb and exchange energy contributions E_J, E_K correspond to the Hartree-Fock energy functional evaluated using the density ρ_{KS} obtained from a semi-local Kohn-Sham (KS) reference determinant

$$E_{\mathrm{HF}}[\{\rho_{\mathrm{KS}}\}] = E_h[\{\rho_{\mathrm{KS}}\}] + E_J[\{\rho_{\mathrm{KS}}\}] + E_K[\{\rho_{\mathrm{KS}}\}]. \quad (2)$$

Using the fluctuation-dissipation theorem and the RPA, the correlation energy can be expressed after coupling-strength integration using the RI as[13,23–26]

$$E_{\mathrm{c}}^{\mathrm{RPA}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\{\log\left(\mathbf{1} - \chi_0(\mathrm{i}\omega)\mathbf{C}\right) + \chi_0(\mathrm{i}\omega)\mathbf{C}\}, \quad (3)$$

where the contravariant Coulomb operator in the auxiliary basis is given as

$$C_{PQ} = (P|m_{12}|R)^{-1}(R|S)(S|m_{12}|Q)^{-1} \quad (4)$$

with the metric in general notation m_{12} chosen for the RI-decomposition. Popular choices for the metric m_{12} are the Coulomb, the overlap, or the attenuated Coulomb metric. Furthermore, $\chi_0(\mathrm{i}\omega)$ is the polarizability of the non-interacting reference system in the imaginary frequency domain in the auxiliary basis. Calculating $\chi_0(\mathrm{i}\omega)$ is done most efficiently by employing a numerical contracted double-Laplace[14] or cosine transform[27] from the imaginary time to the imaginary frequency domain

$$\chi_0(\mathrm{i}\omega) = \int_{-\infty}^{\infty}\mathrm{d}\tau\cos(\omega\tau)\chi_0(\mathrm{i}\tau), \quad (5)$$

$$\chi_0(\mathrm{i}\tau)_{PQ} = B^P_{\mu\nu}G^0_{\mu\lambda}(-\mathrm{i}\tau)G^0_{\nu\sigma}(\mathrm{i}\tau)B^Q_{\lambda\sigma}. \quad (6)$$

Here, $\mathbf{B}$ denotes the three-center ERI tensor

$$B^P_{\mu\nu} = (P|m_{12}|\mu\nu), \quad (7)$$

and the independent particle Green's function $\mathbf{G}^0(\mathrm{i}\tau)$ in imaginary time is given as

$$\mathbf{G}^0(\mathrm{i}\tau) = \Theta(-\mathrm{i}\tau)\underline{\mathbf{G}}^0(\mathrm{i}\tau) + \Theta(\mathrm{i}\tau)\overline{\mathbf{G}}^0(\mathrm{i}\tau) \quad (8)$$

with

$$\underline{G}^0(\mathrm{i}\tau)_{\mu\nu} = \mathcal{C}_{\mu i}\mathcal{C}_{\nu i}e^{-(\varepsilon_i-\varepsilon_F)\tau}, \quad (9)$$

$$\overline{G}^0(\mathrm{i}\tau)_{\mu\nu} = -\mathcal{C}_{\mu a}\mathcal{C}_{\nu a}e^{-(\varepsilon_a-\varepsilon_F)\tau} \quad (10)$$

defined in analogy to the occupied and unoccupied pseudodensities with the MO energies ε_m and MO coefficients $\mathbf{C}$ obtained from the reference KS calculation and the Fermi level ε_F, which ensures numerical stability of the Green's function at large imaginary times. The Heaviside step function is denoted as $\Theta(\mathrm{i}\tau)$.

B. General strategy to obtain first order properties

To obtain first order properties with respect to a perturbation x, we use partial derivatives of the AO-RI-RPA energy with respect to χ_0, $\mathbf{C}$, $\mathbf{G}^0$, and $\mathbf{B}$. The total derivative is then obtained by tracing the partial derivatives with the variation of the respective quantity in response to a perturbation x and summing over all contributions.

This is in similar spirit to the strategy followed by Burow *et al.*[20] To avoid the derivatives of the MO energies ε_m and coefficients $\mathcal{C}_m$, we take, however, a different route as pursued in the canonical RI-RPA gradient theory and calculate derivatives of the non-interacting Green's function in the imaginary time domain as proposed originally for MP2 gradients.[28–30] Along this line, analogies to the plane-wave RPA gradients[21] arise, which extends this general framework for the calculation of first-order properties of electronic structure theories that can be formulated as a functional of $\mathbf{G}^0$ to the AO formalism. To avoid computing the derivative of $\mathbf{G}^0$ with respect to each perturbation x, we take advantage of the Z-vector technique by Handy and Schaefer[31] and its formulation in the AO basis.[28] The derivative of numerical integration roots and weights used for the numerical imaginary time and frequency quadratures which are optimized using the molecular orbital energies will be neglected. Earlier studies on MP2 gradients using the same time grids[28,29] and on RPA gradients using exactly the same time and frequency grids[21] showed that sufficiently accurate results can be obtained without taking these derivatives into account. This is also supported by our results.

We show the derivation in the following for the example of perturbations caused by nuclear displacements in the closed-shell formalism. First, we derive the gradient contribution stemming from the correlation energy and then add the contribution from the Hartree-Fock energy functional evaluated using the KS density. Here caution has to be taken since the Hartree-Fock functional is not stationary with respect to the KS orbitals, which requires to take the response of the density into account as compared to regular Hartree-Fock force calculations, where this can be avoided using the energy weighted one-particle density matrix.[32]

The total gradient can be split up as follows:

$$\frac{\partial E}{\partial x} = E_{\mathrm{HF}}[\{\rho_{\mathrm{KS}}\}]^{(x)} + E_{\mathrm{HF}}[\{\rho^x_{\mathrm{KS}}\}] + E_{\mathrm{c}}^{\mathrm{RPA},x}[\{\mathbf{C},\mathbf{B},\mathbf{G}^0\}], \quad (11)$$

$$E_{\mathrm{c}}^{\mathrm{RPA},x}[\{\mathbf{C},\mathbf{B},\mathbf{G}^0\}] = \mathrm{Tr}\left\{\frac{\partial E_{\mathrm{c}}^{\mathrm{RPA}}}{\partial\mathbf{C}}\frac{\partial\mathbf{C}}{\partial x}\right\} + \mathrm{Tr}\left\{\frac{\partial E_{\mathrm{c}}^{\mathrm{RPA}}}{\partial\mathbf{B}}\frac{\partial\mathbf{B}}{\partial x}\right\} + \mathrm{Tr}\left\{\frac{\partial E_{\mathrm{c}}^{\mathrm{RPA}}}{\partial\mathbf{G}^0}\frac{\partial\mathbf{G}^0}{\partial x}\right\}. \quad (12)$$

The derivatives of the Hartree-Fock functional containing only integral derivatives are denoted as $E_{\mathrm{HF}}[\{\rho_{\mathrm{KS}}\}]^{(x)}$, whereas the contribution stemming from the perturbed density is denoted as $E_{\mathrm{HF}}[\{\rho^x_{\mathrm{KS}}\}]$.

C. Contribution from Coulomb operator in auxiliary basis

First, we derive the expression of the gradient stemming from the Coulomb operator in the auxiliary basis, $\mathbf{C}$ [first term, Eq. (12)]. The contribution to the RI-RPA gradient stemming from $\mathbf{C}$ caused by a nuclear displacement is readily evaluated as

$$\mathrm{Tr}\left\{\frac{\partial E_{\mathrm{c}}^{\mathrm{RPA}}}{\partial\mathbf{C}}\frac{\partial\mathbf{C}}{\partial x}\right\}. \quad (13)$$

The derivative with respect to the Coulomb matrix can be obtained by expanding the logarithm in the RI-RPA energy expression as a series to yield

$$\frac{\partial E_c^{\mathrm{RPA}}}{\partial \mathbf{C}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\{-[(\mathbf{1}-\chi_0(\mathrm{i}\omega)\mathbf{C})^{-1}-\mathbf{1}]\chi_0(\mathrm{i}\omega)\}. \quad (14)$$

The variation of $\mathbf{C}$ with respect to a nuclear displacement x is given as

$$\begin{aligned}\frac{\partial \mathbf{C}}{\partial x} = &-\Pi^{-1}_{m_{12}}\Pi^{x}_{m_{12}}\Pi^{-1}_{m_{12}}\Pi_{r_{12}}\Pi^{-1}_{m_{12}} + \Pi^{-1}_{m_{12}}\Pi^{x}_{r_{12}}\Pi^{-1}_{m_{12}}\\ &-\Pi^{-1}_{m_{12}}\Pi_{r_{12}}\Pi^{-1}_{m_{12}}\Pi^{x}_{m_{12}}\Pi^{-1}_{m_{12}},\end{aligned} \quad (15)$$

using the notation $\Pi_{m_{12},PQ} = (P|m_{12}|Q)$, $\Pi_{r_{12},PQ} = (P|Q)$. The tensors $\Pi^x_{m_{12}}, \Pi^x_{r_{12}}$ contain conventional ERI derivatives with respect to nuclear coordinates for different Coulomb type operators, which can be computed routinely. With $\chi_0(\mathrm{i}\omega)$ being available in a linear-scaling manner from RPA correlation energy calculations,[14–16] no new computationally involved steps appear in this part of the gradient routine.

D. Contribution from three-center integrals

The contribution from three-center integrals [second term, Eq. (12)] to the nuclear gradients is evaluated using

$$\mathrm{Tr}\left\{\frac{\partial E_c^{\mathrm{RPA}}}{\partial \chi_0(\mathrm{i}\omega)}\frac{\partial \chi_0(\mathrm{i}\omega)}{\partial \mathbf{B}}\frac{\partial \mathbf{B}}{\partial x}\right\}, \quad (16)$$

where the trace implies integration over the imaginary frequency axis. The first derivative in Eq. (16) can be evaluated analogously to Eq. (14) and yields for a specific imaginary frequency $\mathrm{i}\omega$

$$\frac{\partial E_c^{\mathrm{RPA}}}{\partial \chi_0(\mathrm{i}\omega)} = -\frac{1}{4\pi}\{\mathbf{C}[(\mathbf{1}-\chi_0(\mathrm{i}\omega)\mathbf{C})^{-1}-\mathbf{1}]\} = -\frac{1}{4\pi}\mathbf{W}_c(\mathrm{i}\omega), \quad (17)$$

as previously shown by Kresse and coworkers.[21] Here, $\mathbf{W}_c(\mathrm{i}\omega)$ is the correlated screened-Coulomb interaction represented in the auxiliary basis set, which is an even function in $\mathrm{i}\omega$-space. Therefore, the contribution of the three-center integrals [Eq. (16)] can be calculated using the imaginary time representation of χ_0 [Eq. (6)] as

$$\begin{aligned}\mathrm{Tr}\Big\{ &-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{d}\tau\cos(\omega\tau)W_c^{PQ}(\mathrm{i}\omega)G^0(-\mathrm{i}\tau)_{\mu\lambda}\\ &\times B^Q_{\lambda\sigma}G^0(\mathrm{i}\tau)_{\sigma\nu}\left(\frac{\partial B}{\partial x}\right)^P_{\mu\nu}\Big\}\end{aligned} \quad (18)$$

$$= \mathrm{Tr}\Big\{-\int_{-\infty}^{\infty}\mathrm{d}\tau W_c^{PQ}(\mathrm{i}\tau)G^0(-\mathrm{i}\tau)_{\mu\lambda}B^Q_{\lambda\sigma}G^0(\mathrm{i}\tau)_{\sigma\nu}\left(\frac{\partial B}{\partial x}\right)^P_{\mu\nu}\Big\}, \quad (19)$$

with the derivative of the three-center RI-integrals $(\frac{\partial B}{\partial x})^P_{\mu\nu}$.

To evaluate Eq. (19) efficiently, we precontract all unperturbed quantities

$$M^P_{\mu\nu} = -2\int_0^{\infty}\mathrm{d}\tau W_c^{PQ}(\mathrm{i}\tau)G^0(-\mathrm{i}\tau)_{\mu\lambda}B^Q_{\lambda\sigma}G^0(\mathrm{i}\tau)_{\sigma\nu}, \quad (20)$$

where the most compute intensive step is the transformation of the three-center integrals with the screened-Coulomb interaction, which has to be performed for each imaginary time grid point. This shows an $\mathcal{O}(N^2_{\mathrm{Aux}}N^2_{\mathrm{Bas}})$ formal scaling behavior, where N_{Aux} and N_{Bas} denote the dimension of the auxiliary and atomic basis sets, respectively. Since $\mathbf{G}^0$ and $\mathbf{B}$ become sparse quantities for large systems when employing a local RI-metric for $\mathbf{B}$,[14] while $\mathbf{W}_c$ is in general a dense matrix, the asymptotic scaling behavior is expected to be quadratic with the molecule size. Both calculating the three-center ERI derivatives and the contraction with $\mathbf{M}$ are not time determining.

E. Contribution from non-interacting Green's function

As the last contribution to the gradient stemming from the correlation energy, the part depending on the perturbed non-interacting Green's function [third term Eq. (12)] is given by

$$\mathrm{Tr}\left\{\frac{\partial E_c^{\mathrm{RPA}}}{\partial \chi_0(\mathrm{i}\omega)}\frac{\partial \chi_0(\mathrm{i}\omega)}{\partial \mathbf{G}^0(\mathrm{i}\tau)}\frac{\partial \mathbf{G}^0(\mathrm{i}\tau)}{\partial x}\right\}, \quad (21)$$

where again the trace implies integration over imaginary frequency and imaginary time. Using the derivative with respect to $\chi_0(\mathrm{i}\omega)$ [Eq. (17)] and the imaginary time representation of χ_0 [Eq. (6)] to calculate the derivative with respect to $G^0(\mathrm{i}\tau)$ in analogy to Eq. (19), we obtain for Eq. (21)

$$\begin{aligned}\mathrm{Tr}\Big\{ &-\frac{1}{2}\int_{-\infty}^{\infty}\mathrm{d}\tau\Big[W_c^{PQ}(\mathrm{i}\tau)B^P_{\mu\lambda}G^0(-\mathrm{i}\tau)_{\lambda\sigma}B^Q_{\sigma\nu}\left(\frac{\partial G^0(\mathrm{i}\tau)}{\partial x}\right)_{\mu\nu}\\ &+W_c^{PQ}(\mathrm{i}\tau)B^P_{\mu\lambda}G^0(\mathrm{i}\tau)_{\lambda\sigma}B^Q_{\sigma\nu}\left(\frac{\partial G^0(-\mathrm{i}\tau)}{\partial x}\right)_{\mu\nu}\Big]\Big\}.\end{aligned} \quad (22)$$

As the correlated self-energy in the GW-approximation in imaginary time is defined as[21,33,34]

$$\Sigma(\mathrm{i}\tau)_{\mu\nu} = -B^P_{\mu\lambda}G^0(\mathrm{i}\tau)_{\lambda\sigma}W_c^{PQ}(\mathrm{i}\tau)B^Q_{\sigma\nu}, \quad (23)$$

we can simplify the above expression

$$\mathrm{Tr}\left\{\frac{1}{2}\int_{-\infty}^{\infty}\mathrm{d}\tau\left[\Sigma(-\mathrm{i}\tau)\left(\frac{\partial \mathbf{G}^0(\mathrm{i}\tau)}{\partial x}\right)+\Sigma(\mathrm{i}\tau)\left(\frac{\partial \mathbf{G}^0(-\mathrm{i}\tau)}{\partial x}\right)\right]\right\} \quad (24)$$

$$= \mathrm{Tr}\left\{\int_{-\infty}^{\infty}\mathrm{d}\tau\Sigma(-\mathrm{i}\tau)\left(\frac{\partial \mathbf{G}^0(\mathrm{i}\tau)}{\partial x}\right)\right\}, \quad (25)$$

where we have used that $\mathbf{W}_c(\mathrm{i}\tau)$ is symmetric in imaginary time.

While Kresse and co-workers[21] proposed a scheme to calculate the derivative of $\mathbf{G}^0$ in the imaginary frequency domain, we employ a method presented by some of us earlier in the context of MP2 gradient theory for the equivalent pseudo-densities.[28–30] Therefore, we employ the series expansion of the matrix exponential in the non-interacting Green's function [Eqs. (9) and (10)] in the AO basis,[28,35,36] e.g., for negative imaginary times

$$\underline{\mathbf{G}}^0(\mathrm{i}\tau) = \mathbf{P}e^{-\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}} = \mathbf{P}\sum_{k=0}^{\infty}\frac{1}{k!}[-\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}]^k, \quad (26)$$

in combination with the series expansion of the derivative of the matrix exponential

$$\frac{\partial \underline{\mathbf{G}}^0(\mathrm{i}\tau)}{\partial x} = \frac{\partial \mathbf{P}}{\partial x} e^{-\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}} + \mathbf{P}\frac{\partial\, e^{-\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}}}{\partial x}, \tag{27}$$

$$\begin{aligned}\frac{\partial\, e^{-\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}}}{\partial x} &= \sum_{k=1}^{\infty}\sum_{l=0}^{k-1}\frac{(-\tau)^k}{k!}[(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}]^l \\ &\quad\times\frac{\partial(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}}{\partial x}[(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}]^{k-l-1},\end{aligned} \tag{28}$$

to relate the derivative of the non-interacting Green's function to the derivatives of the Hamiltonian corresponding to the semi-local KS reference

$$\mathbf{H}_{\mathrm{KS}} = \mathbf{h} + \mathbf{J}[\mathbf{P}] + \mathbf{V}_{\mathrm{KS}}[\mathbf{P}] \tag{29}$$

and the occupied one-particle density matrix

$$P_{\mu\nu} = \lim_{\mathrm{i}\tau\to 0-} G^0(\mathrm{i}\tau)_{\mu\nu} = \mathcal{C}_{\mu i}\mathcal{C}_{\nu i}, \tag{30}$$

calculated from the KS-MO coefficients, as well as the AO overlap matrix $\mathbf{S}$. The notation in Eq. (29) implies

$$J_{\mu\nu}[\mathbf{X}] = (\mu\nu|\lambda\sigma)X_{\lambda\sigma}, \tag{31}$$

$$V_{\mathrm{KS},\mu\nu}[\mathbf{X}] = f^{XC}_{\mu\nu\lambda\sigma}X_{\lambda\sigma}, \tag{32}$$

where $f^{XC}_{\mu\nu\lambda\sigma}$ is the exchange-correlation kernel of the KS reference system in the AO basis. The one-electron Hamiltonian $\mathbf{h}$ contains the kinetic energy and nuclear potential matrices. The same can be performed analogously for positive imaginary times using the virtual one-particle density matrix

$$\overline{\mathbf{G}}^0(\mathrm{i}\tau) = -\mathbf{Q}e^{-\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{Q}}, \tag{33}$$

$$Q_{\mu\nu} = -\lim_{\mathrm{i}\tau\to 0+} G^0(\mathrm{i}\tau)_{\mu\nu} = \mathcal{C}_{\mu a}\mathcal{C}_{\nu a}. \tag{34}$$

To avoid computation of the derivative of the virtual density matrix,[28] the completeness relation

$$\mathbf{1} = \mathbf{PS} + \mathbf{QS}, \tag{35}$$

$$\mathbf{Q}^x = -\mathbf{P}^x - \mathbf{S}^{-1}\mathbf{S}^x\mathbf{S}^{-1} \tag{36}$$

is employed. The details of this procedure can be found in Ref. 28.

The final expression for the gradient contribution stemming from $\mathbf{G}^0$ [Eq. (21)] reads

$$\mathrm{Tr}\{\mathbf{P}_{\mathrm{RPA}}(\mathbf{H}^{(x)}_{\mathrm{KS}} - \varepsilon_F\mathbf{S}^x) + \mathbf{V}_{\mathrm{RPA}}\mathbf{P}^x + \mathbf{S}^{-1}(\overline{\Sigma}^+)\mathbf{S}^{-1}\mathbf{S}^x\}. \tag{37}$$

Here we have defined $\mathbf{V}_{\mathrm{RPA}}$

$$\mathbf{V}_{\mathrm{RPA}} = \overline{\Sigma} + \mathbf{J}[\mathbf{P}_{\mathrm{RPA}}] + \mathbf{V}_{\mathrm{KS}}[\mathbf{P}_{\mathrm{RPA}}], \tag{38}$$

$$\mathbf{P}_{\mathrm{RPA}} = \int_0^\infty \mathrm{d}\tau\,[\mathbf{P}\mathbf{Y}(-\mathrm{i}\tau) - \mathbf{Q}\mathbf{Y}(\mathrm{i}\tau)], \tag{39}$$

$$\overline{\Sigma} = \overline{\Sigma}^- + \overline{\Sigma}^+, \tag{40}$$

$$\overline{\Sigma}^+ = \int_0^\infty \mathrm{d}\tau\,[\mathbf{Y}(\mathrm{i}\tau)(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S}) + e^{-\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{Q}}\Sigma(-\mathrm{i}\tau)], \tag{41}$$

$$\overline{\Sigma}^- = \int_0^\infty \mathrm{d}\tau\,[\mathbf{Y}(-\mathrm{i}\tau)(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S}) + e^{\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}}\Sigma(\mathrm{i}\tau)], \tag{42}$$

with a RPA density matrix $\mathbf{P}_{\mathrm{RPA}}$ and an "integrated" self-energy term $\overline{\Sigma}$ constituent of a negative imaginary time $\overline{\Sigma}^-$ and a positive imaginary time $\overline{\Sigma}^+$ contribution. The auxiliary matrix $\mathbf{Y}(\mathrm{i}\tau)$ is defined as

$$\mathbf{Y}(\mathrm{i}\tau) = \Theta(-\mathrm{i}\tau)\underline{\mathbf{Y}}(\mathrm{i}\tau) + \Theta(\mathrm{i}\tau)\overline{\mathbf{Y}}(\mathrm{i}\tau), \tag{43}$$

$$\begin{aligned}\underline{\mathbf{Y}}(\mathrm{i}\tau) &= \sum_{k=1}^{\infty}\sum_{l=0}^{k-1}\frac{(-\tau)^k}{k!}[(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}]^{k-l-1}\Sigma(-\mathrm{i}\tau)\mathbf{P} \\ &\quad\times[(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}]^l,\end{aligned} \tag{44}$$

$$\begin{aligned}\overline{\mathbf{Y}}(\mathrm{i}\tau) &= \sum_{k=1}^{\infty}\sum_{l=0}^{k-1}\frac{(-\tau)^k}{k!}[(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{Q}]^{k-l-1}\Sigma(-\mathrm{i}\tau)\mathbf{Q} \\ &\quad\times[(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{Q}]^l,\end{aligned} \tag{45}$$

which we calculate recursively as described in Ref. 28.

Furthermore, we note that in Eq. (37) we do not consider the variation caused by the perturbed Fermi level which, as indicated by our results in Sec. IV B, does not seem to have a significant influence. The integral derivatives of the KS Hamiltonian, which contain only basis function derivatives, are denoted as

$$\mathbf{H}^{(x)}_{\mathrm{KS}} = \mathbf{h}^x + \mathbf{J}^x[\mathbf{P}] + \mathbf{V}^x_{\mathrm{KS}}[\mathbf{P}]. \tag{46}$$

In terms of computational complexity, the most involved step is the calculation of the self-energy in imaginary time. Here the time-determining step is again the transformation of the three-center integrals with the correlated screened-Coulomb interaction, which we have discussed in Sec. II D and which serves as an intermediate result for both $\mathbf{M}$ and $\Sigma(\mathrm{i}\tau)$. Here we want to note that a similar efficient calculation of the diagonal elements of the self-energy in the MO basis with a local RI-metric and an AO formalism has been part of previous studies on efficient GW calculations.[34] The calculation of the matrix exponentials and the $\mathbf{Y}(\mathrm{i}\tau)$ matrix is not relevant in terms of computational resources, since it only involves low-prefactor dense linear algebra. To assure fast convergence of all matrix exponentials, the scaling and squaring technique (see, e.g., Ref. 37) is used as described in Ref. 28. The efficient calculation of the perturbed occupied one-particle density matrix $\mathbf{P}^x$ with help of the Z-vector technique is discussed below.

F. Contributions from Hartree-Fock functional

Having dealt with the correlated part of the gradient, the last contribution that remains is the Hartree-Fock functional evaluated with the KS density

$$\frac{\partial E_{\mathrm{HF}}}{\partial x} = E_{\mathrm{HF}}[\{\rho_{\mathrm{KS}}\}]^{(x)} + E_{\mathrm{HF}}[\{\rho^x_{\mathrm{KS}}\}], \tag{47}$$

which we split in the contribution stemming from only integral derivatives and the contribution stemming from the response of the KS density. As mentioned earlier, the latter contribution does not arise at regular self-consistent-field (SCF) level force calculations since the energy functional is then stationary with respect to variations of the density. The explicit form of the first contribution reads

$$E_{\mathrm{HF}}[\{\rho_{\mathrm{KS}}\}]^{(x)} = \mathrm{Tr}\left\{\mathbf{P}\mathbf{h}^x + \frac{1}{2}\left(\mathbf{P}\mathbf{J}^x[\mathbf{P}] + \mathbf{P}\mathbf{K}^x[\mathbf{P}]\right)\right\}, \quad (48)$$

where

$$K^x[\mathbf{P}]_{\mu\nu} = \frac{1}{2}(\mu\lambda|\nu\sigma)^x P_{\lambda\sigma} \quad (49)$$

which can be routinely calculated using SCF force routines. The second part reads

$$E_{\mathrm{HF}}[\{\rho^x_{\mathrm{KS}}\}] = \mathrm{Tr}\left\{\mathbf{P}^x\mathbf{h} + \left(\mathbf{P}^x\mathbf{J}[\mathbf{P}] + \mathbf{P}^x\mathbf{K}[\mathbf{P}]\right)\right\} = \mathrm{Tr}\left\{\mathbf{H}_{\mathrm{HF}}\mathbf{P}^x\right\}, \quad (50)$$

where we have to solve for the derivative of the occupied one-particle density matrix using the coupled-perturbed KS (CPKS) equations for the semi-local KS-functional used to obtain the reference determinant.

G. AO-based Z-Vector equation for the calculation of $\mathbf{P}^x$

Examining Eq. (37) and Eq. (50), one could assume that the solution of the CPKS equation is necessary for each perturbation, which would be computationally costly. One can, however, collect all terms that are contracted with the occupied perturbed one-particle density matrix in the form of

$$\mathrm{Tr}\{(\mathbf{H}_{\mathrm{HF}} + \mathbf{V}_{\mathrm{RPA}})\mathbf{P}^x\}. \quad (51)$$

This form is now amenable to the Z-vector technique,[31] which allows us to only solve one CPKS equation instead of one for each perturbation. The adaption of the Z-vector technique to the AO formalism can be found in Ref. 28. Here we use the density-matrix based Laplace variant of the CPKS equations (DL-CPSCF/DL-CPKS),[38] allowing for an efficient linear-scaling solution of the CPKS equations. Another viable route would be to use a D-CPSCF/D-CPKS method which does not rely on a Laplace transform as, e.g., described in Ref. 39.

H. Rank reduction through Cholesky decomposition

As mentioned earlier, for large molecules, the computationally most expensive linear algebra steps in the calculation of RPA gradients appear in the evaluation of the self-energy [Eq. (23)] and the quantity to be contracted with the three-center integral derivatives $\mathbf{M}$ [Eq. (20)]. For regular atomic and auxiliary basis sets, the time determining step will formally scale as $\mathcal{O}\{N^2_{\mathrm{Aux}}N^2_{\mathrm{Bas}}\}$, similar to the scaling of AO-RI-RPA energy calculations. One disadvantage of AO theories is that while AO quantities show local behavior that allow for low-scaling algorithms, there are significant redundancies in conventional gaussian atomic basis sets, which result in an overhead as compared to canonical theories. Fortunately, these redundancies can oftentimes be removed using rank reduction through Cholesky decomposition (CD)[40,41] with complete pivoting of positive semi-definite matrices. For applications of CD in electronic structure theory, see, e.g., Refs. 15, 29, and 42–46. Pivoted CD is particularly suitable as a matrix decomposition in conjunction with AO methods since it conserves matrix sparsity.[45] The most significant savings can be obtained by pivoted CD of quantities corresponding to occupied orbitals, such as the one-particle density matrix $\mathbf{P}$. We introduce CD by realising that the projection of the non-interacting Green's function on the occupied space keeps the Green's function invariant at negative imaginary times

$$\underline{\mathbf{G}}^0(\mathrm{i}\tau) = \underline{\mathbf{G}}^0(\mathrm{i}\tau)\mathbf{S}\mathbf{P}. \quad (52)$$

Now the one-particle density matrix is a positive semi-definite matrix, which can be Cholesky decomposed as

$$\mathbf{P} = \mathbf{L}\mathbf{L}^T, \quad (53)$$

where $\mathbf{L}$ is a rectangular matrix with N_{Bas} rows and N_{occ} columns, where N_{occ} denotes the number of electrons. Now we can reduce the formal scaling behavior of the calculation of $\mathbf{M}$ by using

$$M^P_{\mu\nu} = -2\int_0^\infty \mathrm{d}\tau W^{PQ}_{\mathrm{c}}(\mathrm{i}\tau)(\mathbf{G}^0(-\mathrm{i}\tau)\mathbf{S}\mathbf{L})_{\mu\underline{i}}(\mathbf{L}^T\mathbf{B}^Q)_{\underline{i}\sigma}G^0(\mathrm{i}\tau)_{\sigma\nu} \quad (54)$$

and contracting the screened-Coulomb interaction with the three-center RI-integral tensor transformed into the Cholesky decomposed density atomic orbital (CDD-AO) basis first, which yields steps that formally scale at most as $\mathcal{O}\{N^2_{\mathrm{Aux}}N_{\mathrm{Bas}}N_{\mathrm{occ}}\}$ just as in CDD-RI-RPA.[15,16] As noted earlier in the asymptotic limit of sparse $\mathbf{G}^0$, $\mathbf{L}$, and $\mathbf{B}$, this time determining step shows quadratic scaling behavior with the molecule size N. In the case where the Green's function and the density matrix are dense and only the sparsity of the three-center integrals is exploited, the asymptotic scaling behavior is cubic. The same idea can be applied for the calculation of the self-energy at negative imaginary times [see Eq. (23)]. More care has to be taken when calculating the self-energy at positive imaginary times, since in this case one can only exploit the rank deficiency of the virtual one-particle density matrix, which does not yield significant computational savings. Upon closer inspection, one realizes, however, that it is possible to calculate

$$\Sigma(\mathrm{i}\tau)\mathbf{P}\mathbf{S} = (\Sigma(\mathrm{i}\tau)\mathbf{L})\mathbf{L}^T\mathbf{S}, \quad (55)$$

which is amenable to the scheme presented above, instead of the plain self-energy at imaginary positive times. Inserting Eq. (55) into the calculation of $\underline{\mathbf{Y}}(\mathrm{i}\tau)$ [Eq. (44)] yields the exact same results due to the idempotency of the one-particle density matrix $\mathbf{P}$. The other place where the self-energy at imaginary positive times appears is in the contraction in the trace with the perturbed density matrix [Eqs. (37), (38), (40), and (42)], where it is precontracted with the matrix exponential $e^{\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}}$. Here it is insightful to look at the different projections of the perturbed density matrix that one solves for. The occupied-virtual and virtual-occupied projections of $\mathbf{P}^x$ are determined using the respective Z-vector projections $\mathbf{Z}_{ov}$ and $\mathbf{Z}_{vo}$. As described in Ref. 38, these are of the form

$$\mathbf{Z}_{ov} = \sum_\tau t_\tau \mathbf{G}^0(-\mathrm{i}\tau)\mathbf{X}\mathbf{G}^0(\mathrm{i}\tau), \quad (56)$$

$$\mathbf{Z}_{vo} = \sum_\tau t_\tau \mathbf{G}^0(\mathrm{i}\tau)\mathbf{X}\mathbf{G}^0(-\mathrm{i}\tau) \quad (57)$$

with the numerical integration over Laplace points. This shows in combination with the series expansion of the matrix-exponential $e^{\tau(\mathbf{H}_{\mathrm{KS}}-\varepsilon_F\mathbf{S})\mathbf{P}}$ [Eq. (26)] that one can insert $\Sigma(\mathrm{i}\tau)\mathbf{P}\mathbf{S}$ or $\mathbf{S}\mathbf{P}\Sigma(\mathrm{i}\tau)$ for the calculation of $\mathbf{Z}_{ov}$ and $\mathbf{Z}_{vo}$, respec-

tively, exploiting again the projection of the non-interacting Green's function at negative imaginary times onto the occupied subspace and the idempotency of **P**. Finally, the occupied-occupied part of the perturbed density matrix can be calculated non-iteratively and without the Z-vector technique as

$$\mathbf{P}^x_{oo} = -\mathbf{P}\mathbf{S}^x\mathbf{P}, \tag{58}$$

where it is obvious that $\Sigma(\mathrm{i}\tau)\mathbf{PS}$ also yields the same result as just the self-energy at positive imaginary times. Therefore CDD enables time determining steps that only scale as $\mathcal{O}(N^2_{\mathrm{Aux}}N_{\mathrm{Bas}}N_{\mathrm{occ}})$, just as in the most efficient CDD-RI-RPA algorithm.[15,16] This means that there is no significant overhead to canonical theories and that large basis sets can be employed, which is prohibitive in pure AO theories.

I. Notes on similar developments in the literature

As the calculation of the correlated self-energy is necessary to obtain first order properties, we want to note that this quantity also appears in the calculation of GW quasiparticle energies, where low-scaling methods on large computing clusters have been reported[34,47] also using methods from earlier studies on RPA correlation energies.[14,48] In comparison to the calculation of the self-energy in the low-scaling massively parallel GW algorithm,[34] where only the diagonal elements of the self energy in the MO basis are calculated, we have additionally introduced CDDs which reduce the formal scaling of our method by a factor of $N_{\mathrm{Bas}}/N_{\mathrm{occ}}$, resulting in a significantly lower prefactor. We want to note that to obtain GW-quasiparticle energies, it is furthermore necessary to transform the self-energy to the imaginary frequency domain first and then extrapolate them to real frequencies before being able to solve the quasiparticle equations, see, e.g., Refs. 33, 34, and 47.

When comparing our method to the canonical RI-RPA gradients,[20] the formal scaling behavior of the rate determining step is basically identical, our implementation uses, however, only 15-20 grid points, while the implementation in TURBOMOLE typically requires more than 60 frequency points. Since the most compute intensive step has to be performed for each frequency grid point, this results in a further reduction of the computational effort. Furthermore, as shown in Sec. IV, our algorithm allows reducing the observed asymptotic scaling behavior to approximately quadratic as compared to $\mathcal{O}(N^4)$ for sparse electronic structures, i.e., extended systems with a non-vanishing HOMO-LUMO gap.

In comparison to the plane-wave method by Kresse and co-workers,[21] we have to evaluate derivatives of basis functions which is not necessary for plane-waves. The use of atomic basis set is, however, essential for the low-scaling scheme presented in our work. In terms of evaluating the contribution stemming from the non-interacting Green's function, we take a similar route. We evaluate the contribution, however, in the imaginary time domain in contrast to the imaginary frequency domain as performed in the plane-wave scheme. The imaginary time technique has proven useful in the calculation of molecular properties in the AO formalism before.[28–30] Concerning the scaling behavior, the plane-wave algorithm shows cubic scaling, while our implementation shows formal quartic scaling before sparsity in the integral or density matrix type quantities is exploited. This is because the non-interacting polarizability and the self-energy are evaluated on a real-space grid where those steps show at most cubic scaling behavior as opposed to quartic when employing gaussian basis functions. The evaluation on a real-space grid is facilitated by fast Fourier transforms from reciprocal to real space and vice versa, which is not possible for molecular systems without translational symmetry. Furthermore, storing the non-interacting polarizability and the self-energy on a real-space grid with the required accuracy would significantly exceed the storage capacity available on conventional computing nodes making this an impractical approach for large molecular systems.

III. COMPUTATIONAL DETAILS

The RPA nuclear gradients were implemented in the FermiONs++[49–51] package developed in our group. We use minimax imaginary time $\{\mathrm{i}\tau_l\}_{l=1}^{N_g}$ and imaginary frequency grids $\{\mathrm{i}\omega_k\}_{k=1}^{N_g}$, with a set of weights $\{w_k\}_{k=1}^{N_g}$ for the numerical frequency integrations and $\{t_l\}_{l=1}^{N_g}$ for the numerical time integrations. Furthermore, the transformation matrix $\gamma_{k,l}$ is used for numerical cosine transformations from imaginary time to imaginary frequency. For the transformation from imaginary frequency to imaginary time, the matrix $\beta_{l,k}$ is employed, which is chosen to fulfil $\beta\cos(\tau\omega) = (\gamma\cos(\omega\tau))^{-1}$. These grids were first developed for RPA correlation energy and GW-quasiparticle energy calculations.[16,27,33] In general, we employ $N_g = 15$ grid points for the imaginary time and frequency grids as proved to be sufficient for RPA correlation energy calculations.[16,27] Currently, our implementation is limited to 20 grid points as the Remez algorithm requires the use of quadruple-precision floating-point operations for the optimization of larger grids. For sparse algebra operations, we use the blocked compressed sparse row (BCSR) format[39,52] which uses an atomic blocking scheme. We aim for a block size of 50 elements and neglect blocks when both the maximum row and maximum column sums regularized to the full matrix dimension fall below 10^{-6}. For a sketch of the algorithm using sparse algebra and an overview on which quantities are stored in the sparse-algebra format, see the supplementary material. Here sparse algebra is employed for the calculation of the self-energies and the transformed three-center integral tensor **M** [see Eqs. (23), (54), and (55)]. The optimal procedure for pivoted CD of occupied density matrices to preserve matrix sparsity is described in, e.g., Refs. 15 and 45. As a threshold for CD, we choose the LAPACK default which is calculated as the product of the dimension of the matrix, the maximum diagonal element, and the machine precision. The details of the recursive calculation of the $\mathbf{Y}(\mathrm{i}\tau)$ matrices [see Eqs. (44) and (45)] can be found in Ref. 28. With a target Frobenius norm of 0.2 for the scaling and squaring step and a threshold of 10^{-10} for the series expansion of the matrix exponential, it was not necessary to go beyond 8th order for any calculation carried out for this manuscript.

As atomic basis sets, we use the def2-SVP, def2-TZVP, and def2-QZVP family of basis sets[53,54] along with the respective RI basis set.[55,56] For the RI for the low-scaling calculations, we use the attenuated Coulomb metric,[15,57–59] which combines locality of the three-center integral tensors similar to the overlap metric, while retaining the accuracy of the most widely used Coulomb metric. We set the attenuation parameter to $\omega = 0.1$.[15] To obtain the KS-reference orbitals, we use the Perdew-Burke-Ernzerhof exchange-correlation functional.[60] The Fermi-level is set in the middle between the highest occupied and lowest unoccupied MO.

IV. RESULTS AND DISCUSSION

First, we benchmark the convergence of the numerical imaginary frequency and imaginary time grids to find the necessary number of grid points to obtain accurate RPA forces. To test the validity of our approach, we first compare our results for a set of small molecules against the implementation in TURBOMOLE[20] and numerical forces. Then we benchmark the effect of our approximations that allow for computational savings, namely, the attenuated RI-metric and the use of sparse algebra, before we show the efficiency of our method and end with an illustrative application. All calculations shown here use the RPA gradient obtained with CDDs since this approach will always be more efficient than the plain AO-RPA implementation and there is no loss of accuracy.

A. Convergence of imaginary time and frequency grids

First, we demonstrate the convergence of the imaginary time and frequency grids[16,27,33] of our RPA force method by performing calculations on the methane-benzene and the uracil dimers of the S22 test set[61] and the anthracene and retinal molecules with def2-QZVP basis sets using 10-20 grid points. The S22 molecules represent a typical application for RPA calculations. Anthracene and retinal are included to show that our method also performs well for small gap systems. The results in Table I show that the minimax imaginary time and frequency grids allow converging the forces to below 10^{-6} a.u./bohr in all cases. The small fluctuations in the calculations for the benzene-methane dimer and anthracene for large grids can be explained from the fact that we converge the solution of the DL-CPKS equations to a residual norm of 10^{-6}, which is approximately the same precision as obtained from the imaginary time and frequency grids. Based on these results, we adopt 15 grid points as a default for the calculations in this manuscript.

B. Validation of the implementation

To validate our CDD-RI-RPA approach, we have performed force calculations on the 36 closed-shell molecules of the G2 test set[62] with the def2-QZVP basis set with the respective RI basis set, using the Coulomb RI-metric and dense algebra. As a comparison, we have performed the same calculations with the canonical RI-RPA implementation in TURBOMOLE[20] and evaluated the forces numerically using the canonical RI-RPA implementation[13] in FermiONs++, however, with the minimax imaginary frequency grid.[27] For the numerical differentiation, we use the five-point stencil method with a step size of 10^{-5} Å. Detailed results for every system can be found in the supplementary material.

TABLE I. Results showing the convergence of RPA force calculations by presenting mean-absolute-deviations (MADs) of the gradient vector calculated with the respective number of grid points against the calculation using $N_g = 20$ grid points. All calculations employed the def2-QZVP basis set using the Coulomb RI-metric and dense algebra.

	Benzene-methane dimer	Uracil-dimer	Anthracene	Retinal
N_g	MAD [10^{-6} a.u./bohr]	MAD [10^{-6} a.u./bohr]	MAD [10^{-6} a.u./bohr]	MAD [10^{-6} a.u./bohr]
10	1.63	7.06	2.89	12.93
11	1.47	4.06	1.62	6.28
12	0.67	2.87	0.48	4.44
13	0.33	1.83	0.42	3.06
14	0.30	1.29	0.33	2.40
15	0.27	0.86	0.29	1.48
16	0.21	0.57	0.23	1.05
17	0.29	0.33	0.28	0.69
18	0.23	0.19	0.30	0.51
19	0.36	0.13	0.27	0.43
20	...	...	...	...

The comparison against numerical gradients (Table II) shows average root-mean-squared deviation (RMSD) and mean-absolute deviation (MAD) on the order of 10^{-6}–10^{-5} a.u./bohr and no maximum error (MAX) larger than 10^{-3} a.u./bohr. The same holds true when comparing against the canonical implementation in TURBOMOLE (Table II), with the average values for RMSD, MAD, and the MAX deviation being slightly higher. This is to be expected since the comparison between two different quantum-chemistry packages brings

TABLE II. Comparison of the results for the 36 closed shell molecules of the G2 test set and the same set of molecules distorted out of equilibrium, obtained with the CDD-RI-RPA approach using the Coulomb RI-metric against numerical forces calculated with the canonical RI-RPA algorithm in FermiONs++ and canonical RI-RPA gradients calculated with TURBOMOLE. We present average root-mean-squared and average mean-absolute deviations of the gradient vector of each system as compared to the respective reference in 10^{-3} a.u./bohr. In addition, we show the maximum deviation for one gradient component of the entire test set. The detailed results for every system can be found in the supplementary material.

	RMSD [10^{-3} a.u./bohr]	MAD [10^{-3} a.u./bohr]	MAX [10^{-3} a.u./bohr]
Numerical forces			
G2	0.008	0.005	0.099
G2 distorted	0.007	0.005	0.079
TURBOMOLE			
G2	0.038	0.025	0.169
G2 distorted	0.036	0.024	0.276

some difficulties as the errors in the calculation of DFT quantities or additional approximations such as the use of the RI make it hard to converge the results to beyond 10^{-5} a.u./bohr precision.

Since the G2 test set is optimized at the MP2/6-31G* level of theory, the absolute values of the gradient vectors are rather small. To also test our implementation at non-equilibrium geometries, we distort one atom in each molecule by 0.2 Å. As shown in Table II, the statistical measures for the distorted molecules basically stay the same showing that our implementation reliably calculates forces accurately also for non-equilibrium geometries.

C. Accuracy of low-scaling CDD-RI-RPA forces

To test the accuracy of our approximations that allow for low-scaling calculation of large molecular systems, we perform calculations using the attenuated RI-metric and sparse algebra for larger molecular systems and again compare against the canonical implementation in TURBOMOLE. In particular, we look at the forces for an amylose chain consisting of four glucose units, a DNA base pair, and a linear alkane chain with 40 carbon atoms using the def2-SVP basis set. We use the attenuated RI-metric (ω = 0.1) and sparse algebra, which allows reducing the computational complexity, as will be shown in Sec. IV D. The results (Table III) show that for large systems we obtain average accuracies of 10^{-5} a.u./bohr and all maximum deviations are smaller than 10^{-3} a.u./bohr. This shows that our approximations do not deteriorate the accuracy of the method.

D. Scaling behavior with molecular size

To test the efficiency of our methods for extended molecules, we performed calculations on linear alkane (up to alkane$_{200}$) and glycine chains (up to glycine$_{60}$) using the def2-SVP basis set. The attenuated RI-metric (ω = 0.1) was used along with sparse algebra. All calculations were performed using 16 threads on a dual-processor Intel Xeon CPU E5-2667 machine. The results for the alkane and glycine chains are shown in Fig. 1. The curve fit through the systems with more than 1900 basis functions using a power law shows

TABLE III. Comparison of the results obtained for three large molecules using the CDD-RI-RPA gradients with the attenuated RI-metric (ω = 0.1) and sparse algebra against the canonical implementation in TURBOMOLE. We have used 15 imaginary time and frequency grid points for the CDD-RI-RPA calculations and as many grid points for the frequency integration in TURBOMOLE necessary to keep the sensitivity measure[20] below 10^{-10}. We present root-mean-squared, mean-absolute, and maximum deviations of the total gradient vector as compared to the respective reference in 10^{-3} a.u./bohr.

	TURBOMOLE		
System	RMSD [10^{-3} a.u./bohr]	MAD [10^{-3} a.u./bohr]	MAX [10^{-3} a.u./bohr]
Amylose$_4$	0.068	0.048	0.336
DNA$_1$	0.086	0.053	0.344
Alkane$_{40}$	0.018	0.009	0.078

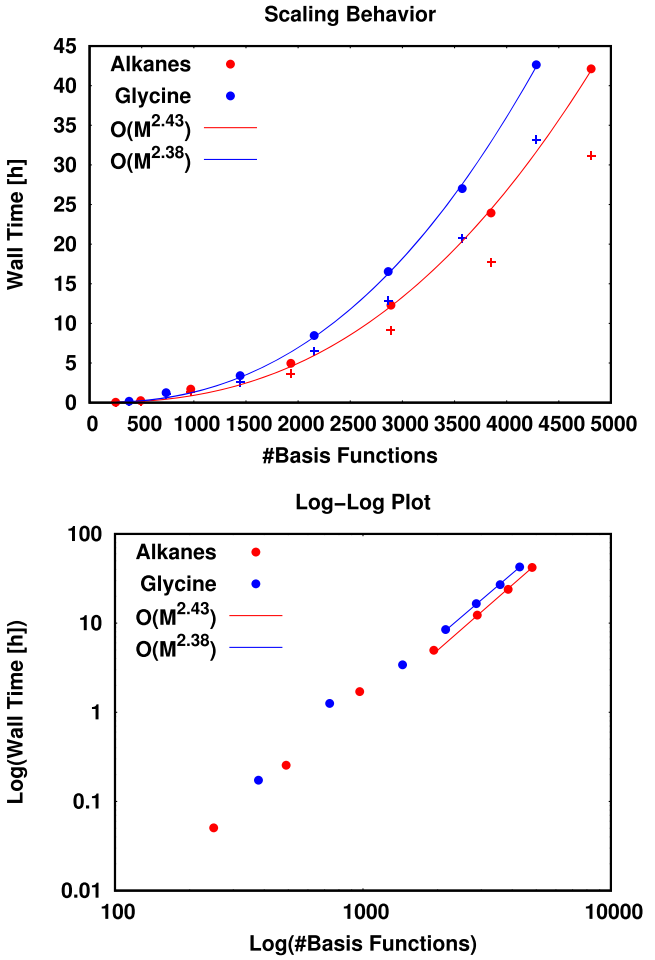


FIG. 1. Timings for the force calculations on a set of linear alkanes and linear glycine chains using the def2-SVP basis set with the corresponding RI basis set. The crosses show the sum of the wall times required for the calculation of the self-energy and the transformed three-center integral tensor. The number of basis functions refers to the number of atomic basis functions. The attenuated RI metric (ω = 0.1) was used along with sparse algebra. The scaling behavior was calculated by fitting a power law to the total wall-times of the systems with more than 1900 basis functions.

that our pilot implementation of the CDD-RI RPA gradients shows asymptotic approximately quadratic scaling behavior with the molecule size. Most of the computational time is spent to compute the three-center integral tensor transformed with the screened-Coulomb matrix as already mentioned in Sec. II. The timings in Fig. 1 confirm that through the use of a local RI-metric and the use of sparse algebra, this step scales approximately quadratically with system size. The reason why the scaling behavior is not perfectly quadratic is presumably the overhead associated with sparse algebra. For large systems, this reduced computational complexity will result in a significant computational benefit as compared to the $\mathcal{O}(N^4)$ scaling canonical RI-RPA gradients.[20]

To show that our method is also applicable to larger basis sets, which are usually required for correlation methods, we performed force calculations with the CDD-RI-RPA method on the set of linear alkanes up to alkane$_{120}$ using the def2-TZVP basis set. The results in Fig. 2 show that even though the asymptotic scaling behavior is higher as compared to the def2-SVP case, it is still sub-cubic and therefore significantly

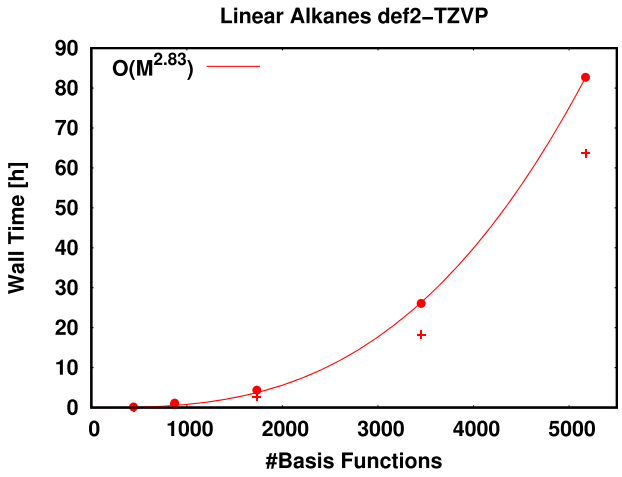


FIG. 2. Timings for the force calculations on a set of linear alkanes using the def2-TZVP basis set with the corresponding RI basis set. The crosses show the sum of the wall times required for the calculation of the self-energy and the transformed three-center integral tensor. The number of basis functions refers to the number of atomic basis functions. The attenuated RI metric ($\omega = 0.1$) was used along with sparse algebra. The scaling behavior was calculated by fitting a power law to the total wall-times of the systems with more than 1700 basis functions.

lower than $\mathcal{O}(N^4)$. This shows that also for larger basis sets the CDD-RI-RPA gradients show better scaling behavior than the canonical method. For quadruple-zeta basis sets, the size of systems available is limited by the large prefactor so that we are not yet entering the regime where we would expect significant speedups with the sparse-algebra implementation. To reduce the basis set requirements, one could resort to range-separated variants of RPA,[63,64] where our gradient scheme is applicable as well.

Finally, we performed calculations on more globular amylose helices ranging from 1 to 16 glucose units using the def2-SVP basis set. As shown in Fig. 3, also for these systems

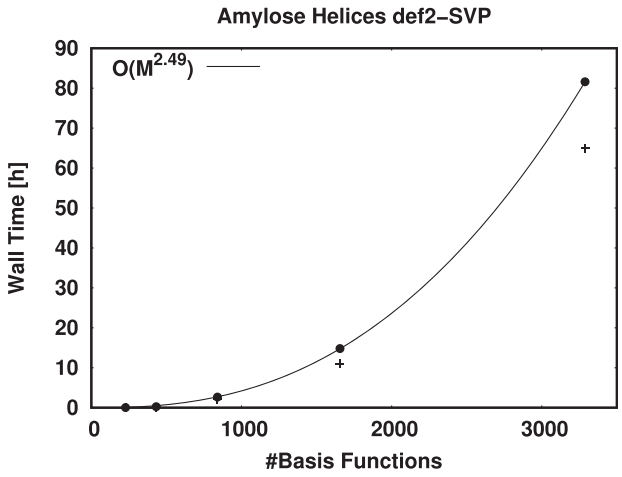


FIG. 3. Timings for the force calculations on a set of amylose helices using the def2-SVP basis set with the corresponding RI basis set. The crosses show the sum of the wall times required for the calculation of the self-energy and the transformed three-center integral tensor. The number of basis functions refers to the number of atomic basis functions. The attenuated RI metric ($\omega = 0.1$) was used along with sparse algebra. The scaling behavior was calculated by fitting a power law to the total wall-times of the systems with more than 800 basis functions.

sub-cubic scaling behavior can be demonstrated. While these systems are not fully three-dimensional, the results indicate that also for non-linear molecules significant computational savings can be obtained with our method.

To show that the calculation of the self energy and the transformed three-center integrals is actually the time determining step for extended molecules, we have also plotted the timings for these steps in Figs. 1–3. For the largest two systems of each plot, this step requires 70%-80% of the total wall time. This shows that for the currently tractable largest systems, this is still the time determining step and not the small-prefactor cubic scaling steps.

E. Scaling behavior with basis set size

A drawback of AO quantities as compared to canonical theories is oftentimes the scaling with respect to the basis set size due to the redundancies present in typical AO basis sets. In our method, we reduce those redundancies through the use of CDDs. The formal scaling behavior of our method is $\mathcal{O}(N^2_{\mathrm{Aux}} N_{\mathrm{Bas}} N_{\mathrm{occ}})$ and should therefore demonstrate cubic scaling with respect to the basis set size at constant molecule size since all quantities except the number of occupied orbitals are basis set dependent. This is the same scaling behavior as is to be expected for canonical-RI- or CDD-RI-RPA energy calculations.[13,15] Therefore our method does not increase the scaling behavior with respect to basis set size as compared to the plain energy calculations. To demonstrate that this is indeed the case we have performed force calculations on three molecules of the L7 test set[65] ranging from 15 to 56 atoms with basis set sizes from double- to quadruple-zeta quality. The results presented in Fig. 4 confirm the cubic scaling with respect to basis set size. As for such rather small systems most time is actually spent to form the Coulomb type matrices with the Z-vector in solving the DL-CPSCF equations, we also show the timings for just the calculation of the self-energy and the transformed three-center integrals which are the computationally most expensive steps for large systems as shown in Sec. IV D. As expected, these operations also show approximately cubic scaling behavior. The reason why these steps show consistently sub-cubic scaling behavior is most likely that linear-algebra routines get more efficient for larger matrix-dimensions as encountered in the calculations with larger basis sets. To speed up the formation of Coulomb type matrices, one could resort to the RI approximation or a GPU accelerated algorithm.[49,66,67]

F. Illustrative calculation

Since RPA total energies improve significantly over plain DFT calculations for dispersion interactions, one potential area of application of our method is structure optimization of large dispersion dominated molecular complexes. Examples of those systems are collected in the L7 test set.[65] As an illustrative calculation, we performed a geometry optimization of the octadecane dimer (112 atoms) of the L7 test set at the RI-RPA level of theory using the def2-TZVP basis set with the corresponding RI basis set. We used the attenuated RI metric along with dense algebra. The structures in the test set were

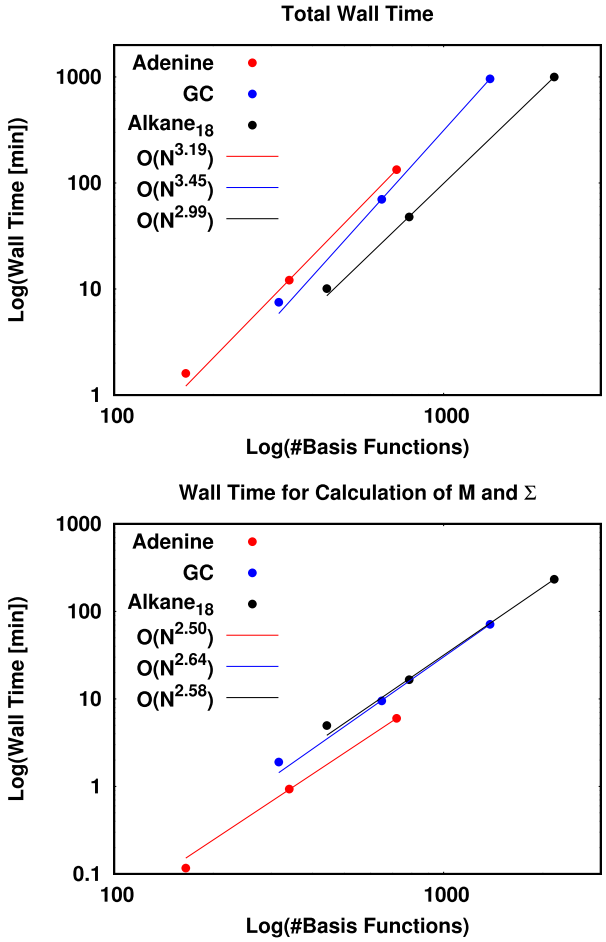


FIG. 4. Timings for the complete force calculation and only for the calculation of the self-energy and the transformed three-center integrals of an adenine molecule, alkane$_{18}$, and a guanine-cytosine base pair (GC) with the def2-SVP, def2-TZVP, and def2-QZVP basis set using dense algebra. The number of basis functions refers to the number of atomic basis functions. The scaling behavior was calculated by fitting a power law.

optimized at the DFT-D TPSS[68] level of theory also employing a triple-zeta basis set, using an empirical dispersion term. Starting from the DFT-D structure, our geometry optimization converged after 5 iterations. As convergence criteria, we used an energy difference criterion of 5×10^{-5} a.u. and a maximum norm of the gradient vector of 5×10^{-4} a.u./bohr. The RMSD of atomic positions of the structure optimized at the DFT-D and the RPA level of theory is 0.4 pm. This shows that the structure obtained using the empirical dispersion correction is in very good agreement with the non-empirical RPA structure. In the future, the CDD-RI-RPA gradients presented in this work could, e.g., serve as a reference to benchmark geometries obtained from dispersion corrected DFT schemes for large molecular complexes.

V. CONCLUSION

In this work, we have derived and implemented analytical gradients for the RPA in an AO framework. Furthermore, we have shown that also for gradients the introduction of CDDs is possible allowing for a reduction of the formal scaling behavior, which is the same as the formal scaling behavior of the CDD-RI-RPA correlation energies. The presented approach was shown to be accurate and asymptotically quadratic scaling, which will allow for significant computational savings for extended molecules with a non-vanishing HOMO-LUMO gap as compared to the $\mathcal{O}(N^4)$ scaling canonical gradients.[20]

The derivation is given in a framework that allows to easily extend the gradient scheme to other methods, which are a functional of the non-interacting Green's function and can be formulated using the RI, such as RPA with exchange methods,[26,69] by simply calculating the respective self-energy and the quantities to be contracted with the perturbed integral tensors, with the rest of the algorithm remaining unchanged. This allows for a straightforward implementation of first order properties that fulfill the abovementioned criteria in the AO framework, similar to the plane-wave scheme presented by Ramberger *et al.*[21]

Furthermore, extending the RPA gradient scheme to other first order properties such as hyperfine coupling constants without spin-orbit couplings is straightforward, see, e.g., our work at the MP2 level.[30] In this case, the perturbation does not affect basis function quantities, i.e., the two- and three-center integral derivatives do not need to be computed.

In the AO-RPA algorithm, the full correlated self-energy in the GW-approximation is calculated. As a side-product, one could obtain all GW-quasiparticle energies at practically no additional computational cost, since the most compute intensive steps are already completed. Furthermore, the low-scaling calculation of the self-energy could be beneficial in self-consistent Green's function calculations in the spirit of second-order Green's function calculations.[70]

SUPPLEMENTARY MATERIAL

See supplementary material for further data as indicated in the text.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Daniel Graf, Dr. Henry Schurkus, and Sigurd Vogler (LMU). Financial support was provided by the Deutsche Forschungsgemeinschaft (DFG) within the Excellence Cluster EXC114 (CIPSM) and the SFB749. C.O. acknowledges, in addition, financial support as a Max-Planck Fellow at the MPI-FKF Stuttgart.

[1] D. Hait and M. Head-Gordon, J. Chem. Theory Comput. **14**, 1969 (2018).
[2] J. Pérez-Jordá and A. Becke, Chem. Phys. Lett. **233**, 134 (1995).
[3] S. Kristyán and P. Pulay, Chem. Phys. Lett. **229**, 175 (1994).
[4] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, Chem. Rev. **116**, 5105 (2016).
[5] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
[6] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).
[7] J. Čížek, J. Paldus, and L. Šroubková, Int. J. Quantum Chem. **3**, 149 (1969).
[8] H. Eshuis, J. E. Bates, and F. Furche, Theor. Chem. Acc. **131**, 1084 (2012).
[9] X. Ren, P. Rinke, C. Joas, and M. Scheffler, J. Mater. Sci. **47**, 7447 (2012).
[10] G. P. Chen, V. K. Voora, M. M. Agee, S. G. Balasubramani, and F. Furche, Annu. Rev. Phys. Chem. **68**, 421 (2017).
[11] J. F. Dobson, J. Wang, B. P. Dinte, K. McLennan, and H. M. Le, Int. J. Quantum Chem. **101**, 579 (2005).
[12] J. F. Dobson and T. Gould, J. Phys. Condens. Matter **24**, 073201 (2012).
[13] H. Eshuis, J. Yarkony, and F. Furche, J. Chem. Phys. **132**, 234114 (2010).
[14] H. F. Schurkus and C. Ochsenfeld, J. Chem. Phys. **144**, 031101 (2016).

[15]A. Luenser, H. F. Schurkus, and C. Ochsenfeld, J. Chem. Theory Comput. **13**, 1647 (2017).
[16]D. Graf, M. Beuerle, H. F. Schurkus, A. Luenser, G. Savasci, and C. Ochsenfeld, J. Chem. Theory Comput. **14**, 2505 (2018).
[17]M. Kállay, J. Chem. Phys. **142**, 204105 (2015).
[18]J. Rekkedal, S. Coriani, M. F. Iozzi, A. M. Teale, T. Helgaker, and T. B. Pedersen, J. Chem. Phys. **139**, 081101 (2013).
[19]B. Mussard, P. G. Szalay, and J. G. Ángyán, J. Chem. Theory Comput. **10**, 1968 (2014).
[20]A. M. Burow, J. E. Bates, F. Furche, and H. Eshuis, J. Chem. Theory Comput. **10**, 180 (2014).
[21]B. Ramberger, T. Schäfer, and G. Kresse, Phys. Rev. Lett. **118**, 106403 (2017).
[22]O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B **13**, 4274 (1976).
[23]D. Langreth and J. Perdew, Solid State Commun. **17**, 1425 (1975).
[24]D. C. Langreth and J. P. Perdew, Phys. Rev. B **15**, 2884 (1977).
[25]J. G. Ángyán, R.-F. Liu, J. Toulouse, and G. Jansen, J. Chem. Theory Comput. **7**, 3116 (2011).
[26]B. Mussard, D. Rocca, G. Jansen, and J. G. Ángyán, J. Chem. Theory Comput. **12**, 2191 (2016).
[27]M. Kaltak, J. Klimeš, and G. Kresse, J. Chem. Theory Comput. **10**, 2498 (2014).
[28]S. Schweizer, B. Doser, and C. Ochsenfeld, J. Chem. Phys. **128**, 154101 (2008).
[29]S. Vogler, M. Ludwig, M. Maurer, and C. Ochsenfeld, J. Chem. Phys. **147**, 024101 (2017).
[30]S. Vogler, G. Savasci, M. Ludwig, and C. Ochsenfeld, J. Chem. Theory Comput. **14**, 3014 (2018).
[31]N. C. Handy and H. F. Schaefer, J. Chem. Phys. **81**, 5031 (1984).
[32]P. Pulay, Mol. Phys. **17**, 197 (1969).
[33]P. Liu, M. Kaltak, J. Klimeš, and G. Kresse, Phys. Rev. B **94**, 165109 (2016).
[34]J. Wilhelm, D. Golze, L. Talirz, J. Hutter, and C. A. Pignedoli, J. Phys. Chem. Lett. **9**, 306 (2018).
[35]P. Y. Ayala and G. E. Scuseria, J. Chem. Phys. **110**, 3660 (1999).
[36]P. R. Surján, Chem. Phys. Lett. **406**, 318 (2005).
[37]T. C. Fung, Int. J. Numer. Methods Eng. **59**, 1273 (2004).
[38]M. Beer and C. Ochsenfeld, J. Chem. Phys. **128**, 221102 (2008).
[39]J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **127**, 054103 (2007).
[40]N. J. Higham, Wiley Interdiscip. Rev.: Comput. Stat. **1**, 251 (2009).
[41]H. Harbrecht, M. Peters, and R. Schneider, Appl. Numer. Math. **62**, 428 (2012).
[42]A. Luenser, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **145**, 124103 (2016).
[43]J. Zienau, L. Clin, B. Doser, and C. Ochsenfeld, J. Chem. Phys. **130**, 204112 (2009).
[44]H. Koch, A. S. de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003).
[45]S. A. Maurer, L. Clin, and C. Ochsenfeld, J. Chem. Phys. **140**, 224112 (2014).
[46]H. F. Schurkus, A. Luenser, and C. Ochsenfeld, J. Chem. Phys. **146**, 211106 (2017).
[47]J. Wilhelm, M. Del Ben, and J. Hutter, J. Chem. Theory Comput. **12**, 3623 (2016).
[48]J. Wilhelm, P. Seewald, M. Del Ben, and J. Hutter, J. Chem. Theory Comput. **12**, 5851 (2016).
[49]J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013).
[50]J. Kussmann and C. Ochsenfeld, J. Chem. Theory Comput. **11**, 918 (2015).
[51]J. Kussmann and C. Ochsenfeld, J. Chem. Theory Comput. **13**, 3153 (2017).
[52]F. G. Gustavson, ACM Trans. Math. Software **4**, 250 (1978).
[53]F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).
[54]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[55]F. Weigend, M. Haser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[56]C. Hättig, Phys. Chem. Chem. Phys. **7**, 59 (2005).
[57]Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, Proc. Natl. Acad. Sci. U. S. A. **102**, 6692 (2005).
[58]Y. Jung, Y. Shao, and M. Head-Gordon, J. Comput. Chem. **28**, 1953 (2007).
[59]S. Reine, E. Tellgren, A. Krapp, T. Kjærgaard, T. Helgaker, B. Jansik, S. Høst, and P. Salek, J. Chem. Phys. **129**, 104101 (2008).
[60]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[61]P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[62]L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).
[63]J. Toulouse, I. C. Gerber, G. Jansen, A. Savin, and J. G. Ángyán, Phys. Rev. Lett. **102**, 096404 (2009).
[64]B. G. Janesko, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **130**, 081105 (2009).
[65]R. Sedlak, T. Janowski, M. Pitonak, J. Rezac, P. Pulay, and P. Hobza, J. Chem. Theory Comput. **9**, 3364 (2013).
[66]I. S. Ufimtsev and T. J. Martínez, J. Chem. Theory Comput. **4**, 222 (2008).
[67]I. S. Ufimtsev and T. J. Martinez, J. Chem. Theory Comput. **5**, 1004 (2009).
[68]P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub, J. Comput. Chem. **28**, 555 (2007).
[69]M. Beuerle, D. Graf, H. F. Schurkus, and C. Ochsenfeld, J. Chem. Phys. **148**, 204104 (2018).
[70]J. J. Phillips and D. Zgid, J. Chem. Phys. **140**, 241101 (2014).

Supplementary Material: Low-Scaling Analytical Gradients for the Direct Random-Phase-Approximation using an Atomic-Orbital Formalism

Matthias Beuerle, Christian Ochsenfeld[a)]

Chair of Theoretical Chemistry,

Department of Chemistry,

University of Munich (LMU),

Butenandtstr. 7, D-81377 München, Germany

[a)] christian.ochsenfeld@uni-muenchen.de

Contents

1 Details on the Calculation of M and $\mathbf{\Sigma}(\mathrm{i}\tau)$ using Sparse Algebra.

Table S1: Summary of the quantities entering the calculation of $\mathbf{M}$ and $\mathbf{\Sigma}(\mathrm{i}\tau)$ with the format of storage and the formal and theoretical asymptotic storage requirements. Sparse matrices are BCSR matrices while all three index tensors are stored as vectors of BCSR matrices.

Quantity	Storage Format	Formal Dimension	Asymptotic Dimension
$\mathbf{W}_\mathrm{c}(\mathrm{i}\tau)$	dense	$N_\mathrm{Aux} \cdot N_\mathrm{Aux}$	$\mathcal{O}(N^2)$
$\mathbf{G}^0(\mathrm{i}\tau)$	sparse	$N_\mathrm{Bas} \cdot N_\mathrm{Bas}$	$\mathcal{O}(N)$
$\mathbf{S}$	sparse	$N_\mathrm{Bas} \cdot N_\mathrm{Bas}$	$\mathcal{O}(N)$
$\mathbf{L}$	sparse	$N_\mathrm{Bas} \cdot N_\mathrm{occ}$	$\mathcal{O}(N)$
$\mathbf{B}$	vec. of sparse	$N_\mathrm{Bas} \cdot N_\mathrm{Bas} \cdot N_\mathrm{Aux}$	$\mathcal{O}(N)$
$\mathbf{BL}$	vec. of sparse	$N_\mathrm{Bas} \cdot N_\mathrm{occ} \cdot N_\mathrm{Aux}$	$\mathcal{O}(N)$

Figure S1: Algorithm showing the sequence of computations carried out to calculate $\mathbf{M}$ and $\mathbf{\Sigma}(\mathrm{i}\tau)$. All quantities that are not listed in Tab. S1 shown in the algorithm are stored as intermediate BCSR matrices. The computationally most demanding step is shown in l.6.

Algorithm 1 Pseudocode for the calculation of the computationally most demanding quantities $\mathbf{M}$ and $\mathbf{\Sigma}(\mathrm{i}\tau)$

1: **for** $P \in N_\mathrm{Aux}$ **do**
2: $\quad [\mathbf{BL}]^P = \mathbf{B}^P\mathbf{L}$
3: **end for**
4: **for** $P \in N_\mathrm{Aux}$ **do**
5: $\quad$ **for** $\tau_n \in N_\tau$ **do**
6: $\quad\quad [\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\mathbf{BL}]^P = W_\mathrm{c}^{PQ}(\mathrm{i}\tau_n)[\mathbf{BL}]^Q$
7: $\quad\quad [\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\overline{\mathbf{G}}(\mathrm{i}\tau_n)\mathbf{BL}]^P = \overline{\mathbf{G}}^0(\mathrm{i}\tau_n)[\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\mathbf{BL}]^P$
8: $\quad\quad [\mathbf{\Sigma}(\mathrm{i}\tau_n)\mathbf{PS}] += \mathbf{B}^P[\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\overline{\mathbf{G}}(\mathrm{i}\tau_n)\mathbf{BL}]^P[\mathbf{L}^T\mathbf{S}]$
9: $\quad\quad \mathbf{\Sigma}(-\mathrm{i}\tau_n) += [\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\mathbf{BL}]^P[\mathbf{L}^T\mathbf{S}\underline{\mathbf{G}}^0(\mathrm{i}\tau_n)]\mathbf{B}^P$
10: $\quad\quad \mathbf{M}^P += [\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\overline{\mathbf{G}}(\mathrm{i}\tau_n)\mathbf{BL}]^P[\mathbf{L}^T\mathbf{S}\underline{\mathbf{G}}^0(\mathrm{i}\tau_n)]$
11: $\quad$ **end for**
12: **end for**

Table S2: Formal and asymptotic complexity of the computationally most demanding linear algebra steps in the caclulation of RPA forces with the corresponding line number of the algorithm sketch in Fig. S1.

Quantity	Line No.	Formal Scaling Behavior	Asymptotic Scaling Behavior
$\mathbf{BL}$	l.2	$\mathcal{O}(N_\mathrm{Aux}N_\mathrm{Bas}^2N_\mathrm{occ})$	$\mathcal{O}(N)$
$\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\mathbf{BL}$	l.6	$\mathcal{O}(N_\mathrm{Aux}^2N_\mathrm{Bas}N_\mathrm{occ})$	$\mathcal{O}(N^2)$
$\mathbf{W}_\mathrm{c}(\mathrm{i}\tau_n)\overline{\mathbf{G}}(\mathrm{i}\tau_n)\mathbf{BL}$	l.7	$\mathcal{O}(N_\mathrm{Aux}N_\mathrm{Bas}^2N_\mathrm{occ})$	$\mathcal{O}(N^2)$
$\mathbf{\Sigma}(\mathrm{i}\tau_n)\mathbf{PS}$	l.8	$\mathcal{O}(N_\mathrm{Aux}N_\mathrm{Bas}^2N_\mathrm{occ})$	$\mathcal{O}(N)$
$\mathbf{\Sigma}(-\mathrm{i}\tau_n)$	l.9	$\mathcal{O}(N_\mathrm{Aux}N_\mathrm{Bas}^2N_\mathrm{occ})$	$\mathcal{O}(N)$
$\mathbf{M}$	l.10	$\mathcal{O}(N_\mathrm{Aux}N_\mathrm{Bas}^2N_\mathrm{occ})$	$\mathcal{O}(N^2)$

2 Validation Data

Table S3: Comparison of the results for the G2 test set obtained with the CDD-RI-RPA approach using the Coulomb RI-metric and dense algebra against numerical forces calculated with the canonical RI-RPA algorithm in FermiONs++. We have used 15 imaginary time and frequency grid points for the CDD-RI-RPA calculations and numerical forces. We present root-mean-squared, mean-absolute, and maximum deviations of the total gradient vector as compared to the respective reference in 10^{-3} a.u./Bohr.

System	RMSD	MAD	MAX	System	RMSD	MAD	MAX
C_2H_2	0.016	0.008	0.038	HF	0.003	0.002	0.005
NH_3	0.002	0.001	0.003	HO-OH	0.006	0.004	0.010
CO_2	0.013	0.006	0.030	H_2S	0.003	0.002	0.004
CS	0.002	0.001	0.003	LiF	0.002	0.001	0.004
CO	0.011	0.007	0.020	LiH	0.001	0.000	0.002
Cl_2	0.009	0.005	0.016	CH_4	0.003	0.002	0.006
ClF	0.009	0.005	0.015	H_3C-SH	0.011	0.007	0.031
Li_2	0.003	0.002	0.006	CH_3-OH	0.004	0.003	0.009
$H_3Si-SiH_3$	0.033	0.019	0.099	CH_3Cl	0.004	0.003	0.009
Na_2	0.001	0.001	0.002	N_2	0.011	0.007	0.020
H_3C-CH_3	0.023	0.013	0.074	P_2	0.002	0.001	0.004
Ethylene	0.018	0.009	0.050	PH_3	0.003	0.002	0.005
F_2	0.006	0.003	0.010	SiH_4	0.009	0.009	0.013
Formaldehyde	0.002	0.001	0.005	SiO	0.009	0.005	0.016
HOCl	0.006	0.004	0.010	CH_2	0.003	0.002	0.005
H_2N-NH_2	0.003	0.003	0.008	SiH_2	0.009	0.007	0.016
HCl	0.007	0.004	0.012	NaCl	0.009	0.005	0.017
HCN	0.017	0.008	0.036	H_2O	0.004	0.003	0.007

Table S4: Comparison of the results obtained for the G2 test set with the CDD-RI-RPA approach using the Coulomb RI-metric and dense algebra against the canonical implementation in TURBOMOLE. We have used 15 imaginary time and frequency grid points for the CDD-RI-RPA calculations and 120 grid points for the frequency integration in the TURBOMOLE calculations. Values labelled with a dagger indicate the use of 240 grid points in the TURBOMOLE calculations. We present root-mean-squared, mean-absolute, and maximum deviations of the total gradient vector as compared to the respective reference in 10^{-3} a.u./Bohr.

System	RMSD	MAD	MAX	System	RMSD	MAD	MAX
C_2H_2	0.015	0.006	0.037	HF	0.056	0.033	0.098
NH_3	0.073	0.057	0.169	HO-OH	0.082	0.069	0.142
CO_2	0.012	0.006	0.027	H_2S	0.055	0.039	0.117
$CS^\dagger$	0.061	0.035	0.106	LiF	0.002	0.001	0.005
CO	0.010	0.006	0.017	LiH	0.007	0.004	0.013
$Cl_2{}^\dagger$	0.050	0.029	0.087	CH_4	0.009	0.008	0.011
$ClF^\dagger$	0.043	0.025	0.076	H_3C-SH	0.032	0.021	0.074
Li_2	0.007	0.004	0.012	CH_3-OH	0.044	0.030	0.108
$H_3Si-SiH_3$	0.037	0.025	0.072	CH_3Cl	0.033	0.021	0.101
$Na_2{}^\dagger$	0.015	0.008	0.025	N_2	0.040	0.023	0.068
H_3C-CH_3	0.021	0.018	0.035	$P_2{}^\dagger$	0.006	0.003	0.010
Ethylene	0.024	0.017	0.043	PH_3	0.017	0.014	0.030
F_2	0.090	0.052	0.157	SiH_4	0.043	0.039	0.048
Formaldehyde	0.020	0.012	0.040	$SiO^\dagger$	0.062	0.036	0.110
$HOCl^\dagger$	0.051	0.034	0.103	$CH_2{}^\dagger$	0.062	0.036	0.150
H_2N-NH_2	0.085	0.063	0.154	$SiH_2{}^\dagger$	0.043	0.031	0.074
HCl	0.060	0.035	0.105	$NaCl^\dagger$	0.005	0.003	0.010
HCN	0.026	0.012	0.056	H_2O	0.071	0.050	0.153

Table S5: Comparison of the results for the distorted G2 test set obtained with the CDD-RI-RPA approach using the Coulomb RI-metric and dense algebra against numerical forces calculated with the canonical RI-RPA algorithm in FermiONs++. We have used 15 imaginary time and frequency grid points for the CDD-RI-RPA calculations and numerical forces. We present root-mean-squared, mean-absolute, and maximum deviations of the total gradient vector as compared to the respective reference in 10^{-3} a.u./Bohr.

System	RMSD	MAD	MAX	System	RMSD	MAD	MAX
C_2H_2	0.017	0.011	0.035	HF	0.004	0.002	0.006
NH_3	0.002	0.001	0.004	HO-OH	0.007	0.006	0.012
CO_2	0.024	0.011	0.055	H_2S	0.005	0.004	0.010
CS	0.002	0.001	0.004	LiF	0.002	0.001	0.004
CO	0.001	0.001	0.002	LiH	0.000	0.000	0.000
Cl_2	0.046	0.026	0.079	CH_4	0.003	0.002	0.006
ClF	0.004	0.002	0.007	H_3C-SH	0.012	0.009	0.031
Li_2	0.001	0.000	0.001	CH_3-OH	0.006	0.005	0.014
$H_3Si-SiH_3$	0.020	0.013	0.063	CH_3Cl	0.007	0.004	0.020
Na_2	0.003	0.001	0.004	N_2	0.002	0.001	0.003
H_3C-CH_3	0.013	0.009	0.038	P_2	0.010	0.005	0.017
Ethylene	0.011	0.007	0.034	PH_3	0.006	0.005	0.012
F_2	0.009	0.005	0.016	SiH_4	0.003	0.003	0.005
Formaldehyde	0.003	0.002	0.007	SiO	0.006	0.003	0.010
HOCl	0.004	0.003	0.008	CH_2	0.004	0.003	0.007
H_2N-NH_2	0.006	0.004	0.013	SiH_2	0.002	0.001	0.004
HCl	0.007	0.004	0.013	NaCl	0.003	0.002	0.005
HCN	0.012	0.008	0.026	H_2O	0.002	0.001	0.004

Table S6: Comparison of the results obtained for the distorted G2 test set with the CDD-RI-RPA approach using the Coulomb RI-metric and dense algebra against the canonical implementation in TURBOMOLE. We have used 15 imaginary time and frequency grid points for the CDD-RI-RPA calculations and 120 grid points for the frequency integration in the TURBOMOLE calculations. Values labelled with a dagger indicate the use of 240 grid points in the TURBOMOLE calculations. We present root-mean-squared, mean-absolute, and maximum deviations of the total gradient vector as compared to the respective reference in 10^{-3} a.u./Bohr.

System	RMSD	MAD	MAX	RMSD	MAD	MAX	
C_2H_2	0.038	0.021	0.083	HF	0.021	0.012	0.037
NH_3	0.053	0.029	0.159	HO-OH	0.130	0.102	0.276
CO_2	0.056	0.029	0.132	$H_2S^\dagger$	0.023	0.013	0.056
$CS^\dagger$	0.118	0.068	0.205	LiF	0.004	0.002	0.008
CO	0.009	0.005	0.016	LiH	0.010	0.006	0.017
$Cl_2^\dagger$	0.061	0.035	0.105	CH_4	0.017	0.013	0.030
$ClF^\dagger$	0.001	0.001	0.003	H_3C-SH	0.037	0.027	0.078
Li_2	0.006	0.004	0.011	CH_3-OH	0.046	0.036	0.100
$H_3Si-SiH_3$	0.040	0.028	0.118	CH_3Cl	0.017	0.008	0.056
$Na_2^\dagger$	0.012	0.007	0.021	N_2	0.005	0.003	0.009
H_3C-CH_3	0.028	0.022	0.071	$P_2^\dagger$	0.087	0.050	0.151
Ethylene	0.023	0.015	0.049	PH_3	0.027	0.021	0.046
$F_2^\dagger$	0.014	0.008	0.023	SiH_4	0.061	0.046	0.153
Formaldehyde†	0.020	0.013	0.043	$SiO^\dagger$	0.002	0.001	0.003
$HOCl^\dagger$	0.043	0.035	0.080	$CH_2^\dagger$	0.023	0.016	0.052
H_2N-NH_2	0.100	0.077	0.211	$SiH_2^\dagger$	0.068	0.047	0.150
$HCl^\dagger$	0.011	0.006	0.019	$NaCl^\dagger$	0.011	0.007	0.020
HCN	0.047	0.023	0.104	H_2O	0.015	0.011	0.031

3.4 Publication IV: Short-range second order screened exchange correction to RPA correlation energies

M. Beuerle, C. Ochsenfeld,
"Short-Range Second Order Screened Exchange Correction to RPA Correlation Energies",
J. Chem. Phys., **147**, 204107 (2017).

Abstract: Direct random phase approximation (RPA) correlation energies have become increasingly popular as a post-Kohn-Sham correction, due to significant improvements over DFT calculations for properties such as long-range dispersion effects, which are problematic in conventional density functional theory. On the other hand, RPA still has various weaknesses, such as unsatisfactory results for non-isogyric processes. This can in parts be attributed to the self-correlation present in RPA correlation energies, leading to significant self-interaction errors. Therefore a variety of schemes have been devised to include exchange in the calculation of RPA correlation energies in order to correct this shortcoming. One of the most popular RPA plus exchange schemes is the second order screened exchange (SOSEX) correction. RPA + SOSEX delivers more accurate absolute correlation energies and also improves upon RPA for non-isogyric processes. On the other hand, RPA + SOSEX barrier heights are worse than those obtained from plain RPA calculations. To combine the benefits of RPA correlation energies and the SOSEX correction, we introduce a short-range RPA + SOSEX correction. Proof of concept calculations and benchmarks showing the advantages of our method are presented.

THE JOURNAL OF CHEMICAL PHYSICS **147**, 204107 (2017)

Short-range second order screened exchange correction to RPA correlation energies

Matthias Beuerle and Christian Ochsenfeld[a)]
Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 München, Germany and Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5–13, D-81377 München, Germany

(Received 2 August 2017; accepted 20 October 2017; published online 29 November 2017)

Direct random phase approximation (RPA) correlation energies have become increasingly popular as a post-Kohn-Sham correction, due to significant improvements over DFT calculations for properties such as long-range dispersion effects, which are problematic in conventional density functional theory. On the other hand, RPA still has various weaknesses, such as unsatisfactory results for non-isogyric processes. This can in parts be attributed to the self-correlation present in RPA correlation energies, leading to significant self-interaction errors. Therefore a variety of schemes have been devised to include exchange in the calculation of RPA correlation energies in order to correct this shortcoming. One of the most popular RPA plus exchange schemes is the second order screened exchange (SOSEX) correction. RPA + SOSEX delivers more accurate absolute correlation energies and also improves upon RPA for non-isogyric processes. On the other hand, RPA + SOSEX barrier heights are worse than those obtained from plain RPA calculations. To combine the benefits of RPA correlation energies and the SOSEX correction, we introduce a short-range RPA + SOSEX correction. Proof of concept calculations and benchmarks showing the advantages of our method are presented. *Published by AIP Publishing.* https://doi.org/10.1063/1.4998647

I. INTRODUCTION

In recent years, the direct random phase approximation (RPA) has become an increasingly popular post-Kohn-Sham method to obtain correlation energies (for an overview over recent developments see, e.g., Refs. 1–3). There exists a variety of formulations of how to calculate RPA correlation energies. Two popular frameworks are the adiabatic connection fluctuation dissipation theorem (ACFDT)[4–8] and the direct ring coupled cluster (drCCD) method.[9] In the ACFDT framework, the correlation energy is obtained utilizing linear response functions obtained at the direct RPA level of theory. This entails neglecting everything but direct (Coulomb) contributions in the calculation of the linear response function. Approaches including either approximate or exact exchange kernels in the calculation of the response function are commonly termed as RPA with exchange (RPAx, see, e.g., Refs. 6 and 10–17).

Equivalently, the direct RPA correlation energy can be calculated by contracting the drCCD amplitudes with two-electron integrals.[9] The drCCD equations can be obtained from the ring coupled cluster (rCCD) method by neglecting all exchange contributions. The rCCD method on the other hand can be derived from the coupled cluster doubles (CCD) method by keeping only particle-hole ring contractions.

The strengths and weaknesses of RPA correlation energies have been extensively discussed in the literature.[7,14,18–24] It has been shown that RPA is good at describing static correlation to some extent, as exemplified by the paradigm system of closed-shell stretched H_2, which is problematic for, e.g., conventional DFT methods.[18,19] In terms of properties, RPA correlation energies have been shown to provide a good description of long-range dispersion interactions,[20,21] which is also one of the big weaknesses of conventional DFT. Furthermore, reaction barrier heights are described very well within RPA, also compared to several RPA correction schemes, which has also been attributed to the good description of static correlation within RPA.[22]

On the other hand, the performance of RPA energies for short-range correlations and interactions and non-isogyric processes such as ionisation potentials and atomisation energies is rather poor.[7,14,22–24] This can in parts be attributed to the self-correlation present in direct RPA energies due to the neglect of all exchange terms, leading to a significant self-interaction error.[18] To circumvent this shortcoming of RPA correlation energies, several approaches have been suggested. Among those methods, RPA including second-order screened exchange (RPA + SOSEX) introduces exchange via contraction of the drCCD amplitudes with antisymmetrised two-electron integrals in the CC framework.[25,26] A similar, however, not completely identical RPA + SOSEX variant contracts the linear response functions with antisymmetrised two-electron integrals in the ACFDT framework.[6,10,27] Other variants of RPA with exchange introduce exchange directly in the calculation of the interacting linear response function or the calculation of the double amplitudes (see, e.g., Refs. 9–17 and 28).

Schemes that are in principle compatible with all RPA variants, are range separated variants of the RPA correlation

[a)]Electronic mail: christian.ochsenfeld@uni-muenchen.de

energy.[16,17,29–33] In these methods, long-range RPA correlation is combined with a short-range density functional description. The purpose of these schemes is to overcome the problems associated with the description of short-range RPA correlation energies while keeping the good performance for long-range dispersion effects. Another advantage that has been mentioned is increased convergence of the correlation energy with respect to the basis set size.[33]

In this manuscript, we suggest to use range separation in the calculation of RPA + SOSEX energies in a different manner. While it has been shown that including SOSEX reduces the self-interaction error of RPA and improves upon RPA in this respect, RPA + SOSEX suffers from a lacking description of static correlation in the long range regime, which, for example, deteriorates the RPA dissociation curve of closed shell H_2.[18] In terms of properties, adding the RPA + SOSEX correction leads to a better description of non-isogyric processes such as ionisation potentials and atomisation energies.[14,22,24,25] At the same time, however, RPA + SOSEX barrier heights are worse than those obtained from a plain RPA calculation.[14,22,24] Those findings can be brought in line with the fact that RPA + SOSEX improves upon the self-interaction error of RPA, while sacrificing the good description of static correlation.[18] With these results in mind, we propose and test a post-Kohn-Sham functional that uses a short-range SOSEX correction in combination with full-range RPA. The idea behind the short-range SOSEX correction is to correct for the self-interaction error present in RPA, which is dominant in the short-range regime, while the good description of RPA for long-range static correlation is preserved.

II. THEORY

A. RPA SOSEX drCCD formalism

The SOSEX correction to RPA correlation energies was first introduced in the drCCD framework.[25,26] Herein, the RPA correlation energy is obtained by contracting the drCCD amplitudes with two-electron integrals

$$E_c^{\mathrm{RPA}} = \frac{1}{2}\mathrm{Tr}\left\{\mathbf{T}_{\mathrm{drCCD}}\mathbf{B}\right\}, \tag{1}$$

where $\mathbf{T}_{\mathrm{drCCD}}$ obeys the Riccati equation[9]

$$\mathbf{B} + \mathbf{AT} + \mathbf{TA} + \mathbf{TBT} = \mathbf{0}, \tag{2}$$

with

$$A_{ia,jb} = \langle ib|aj\rangle + (\varepsilon_a - \varepsilon_i)\delta_{ij}\delta_{ab}, \tag{3}$$

$$B_{ia,jb} = \langle ij|ab\rangle\,. \tag{4}$$

Here ε_a and ε_i correspond to virtual and occupied orbital energies and $\langle pq|rs\rangle$ represents a two-electron integral in the Dirac notation, where $a, b, \ldots$ denote virtual orbitals and $i, j, \ldots$ denote occupied orbitals. The RPA + SOSEX correlation energy is given by replacing $\mathbf{B}$ with $(\mathbf{B} - \mathbf{K})$ in Eq. (1), where $K_{ia,\,jb} = \langle ij|ba\rangle$[25,26] yielding

$$E_c^{\mathrm{SOSEX}} = \frac{1}{2}\mathrm{Tr}\{\mathbf{T}_{\mathrm{drCCD}}(\mathbf{B}-\mathbf{K})\}. \tag{5}$$

This correction is supposed to improve upon the self-interaction error present in RPA correlation energies. This can be seen exemplarily for the case of a one-electron system for which Eq. (5) delivers the correct result $E_c = 0$, while Eq. (1) gives a nonzero correlation energy. However, introducing the SOSEX correction deteriorates the good performance of RPA for bond dissociations, i.e., long-range correlation, where static correlation effects become important.[18] Therefore we propose an attenuated SOSEX correction $(\mathbf{B} - \mathbf{K}^{\mu})$ using short-range integrals,

$$K^{\mu}_{ia,jb} = \langle ij|ba\rangle^{\mu} = \int\int \frac{\varphi_i(r_1)\varphi_b(r_1)\mathrm{erfc}(\mu r_{12})\varphi_j(r_2)\varphi_a(r_2)}{r_{12}}\,\mathrm{d}r_1\,\mathrm{d}r_2, \tag{6}$$

where μ is the parameter controlling the attenuation of the SOSEX correction. We omit complex conjugation since we employ real orbitals in this work. This short-range SOSEX correction then ideally combines the good short-range description of RPA + SOSEX removing self-correlation with the good description of long-range correlation of RPA. We note that this approach is in a similar spirit as compared to the range separated Brueckner CCD theory.[34] The short-range RPA + μ-SOSEX correlation energy is then evaluated as

$$E_c^{\mu-\mathrm{SOSEX}} = \frac{1}{2}\mathrm{Tr}\{\mathbf{T}_{\mathrm{drCCD}}(\mathbf{B}-\mathbf{K}^{\mu})\}. \tag{7}$$

B. RPA SOSEX ACFDT formalism

Alternatively the RPA and RPA + SOSEX correlation energies can be calculated in the ACFDT framework. All quantities in this section are formulated in the particle-hole orbital basis. The starting point for this formulation is the formula for the correlation energy in the ACFDT,[4,5]

$$E_c^{\mathrm{ACFDT}} = -\frac{1}{2}\int_0^1 \mathrm{d}\alpha \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\{(\mathbb{\Pi}_{\alpha}(\mathrm{i}\omega) - \mathbb{\Pi}_0(\mathrm{i}\omega))\mathbb{V}\}. \tag{8}$$

Here $\mathbb{\Pi}_{\alpha}$ corresponds to the frequency dependent, interacting polarization propagator at coupling strength α, $\mathbb{\Pi}_0$ corresponds to the frequency dependent polarization propagator of the non-interacting reference system, and $\mathbb{V} = \begin{pmatrix}\mathbf{B} & \mathbf{B}\\ \mathbf{B} & \mathbf{B}\end{pmatrix}$ corresponds to the Coulomb interaction in the particle-hole basis. Inserting the RPA polarization propagator yields the formula for the RPA correlation energy

$$E_c^{\mathrm{RPA}} = -\frac{1}{2}\int_0^1 \mathrm{d}\alpha \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\{(\mathbb{1} - \alpha\mathbb{\Pi}_0(\mathrm{i}\omega)\mathbb{V})^{-1}\mathbb{\Pi}_0(\mathrm{i}\omega)\mathbb{V} - \mathbb{\Pi}_0(\mathrm{i}\omega)\mathbb{V}\}. \tag{9}$$

In the ACFDT framework, RPA + SOSEX was introduced by substituting $\mathbb{V}$ in Eq. (8) with antisymmetrised integrals $\mathbb{W} = \begin{pmatrix}\mathbf{B}-\mathbf{K} & \mathbf{B}-\mathbf{K}\\ \mathbf{B}-\mathbf{K} & \mathbf{B}-\mathbf{K}\end{pmatrix}$[6,10] also using the RPA polarization propagator

$$E_c^{\mathrm{SOSEX}} = -\frac{1}{2}\int_0^1 \mathrm{d}\alpha \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\{(\mathbb{1} - \alpha\mathbb{\Pi}_0(\mathrm{i}\omega)\mathbb{V})^{-1}\mathbb{\Pi}_0(\mathrm{i}\omega)\mathbb{W} - \mathbb{\Pi}_0(\mathrm{i}\omega)\mathbb{W}\}. \tag{10}$$

Using real orbitals, we follow Ref. 10 and perform a dimension reduction of all matrices appearing in Eqs. (9) and (10) from $2N_{ph} \times 2N_{ph}$ to $N_{ph} \times N_{ph}$, where N_{ph} denotes the dimension of the particle-hole basis. Furthermore, we also perform

the coupling strength integration analytically as described in Ref. 10 leading to

$$E_c^{\mathrm{SOSEX}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\mathrm{Tr}\{\log\left(\mathbf{1}-\mathbf{\Pi}_0(i\omega)\mathbf{B}\right)\mathbf{B}^{-1}\mathbf{W}+\mathbf{\Pi}_0(i\omega)\mathbf{W}\}. \tag{11}$$

For the dimension reduced case, the polarization propagator of the noninteracting reference corresponds to $\mathbf{\Pi}_0(i\omega)_{ia,jb} = -2\varepsilon_{ia}\delta_{ij}\delta_{ab}(\varepsilon_{ia}^2+\omega^2)^{-1}$ and $\mathbf{W} = \mathbf{B} - \mathbf{K}$. Here we also introduce the short-range SOSEX correction by substituting $\mathbf{K}$ with $\mathbf{K}^\mu$. The short-range RPA + μ-SOSEX correlation energy is then evaluated as follows using the ACFDT:

$$E_c^{\mu-\mathrm{SOSEX}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\mathrm{Tr}\{\log\left(\mathbf{1}-\mathbf{\Pi}_0(i\omega)\mathbf{B}\right)\mathbf{B}^{-1}\mathbf{W}^\mu + \mathbf{\Pi}_0(i\omega)\mathbf{W}^\mu\}, \tag{12}$$

with $\mathbf{W}^\mu = \mathbf{B} - \mathbf{K}^\mu$.

For plain RPA correlation energies, the results obtained from the drCCD and the ACFDT calculations have been shown to be strictly equivalent.[9] The drCCD and ACFDT RPA + SOSEX expressions agree, however, only up to second order of perturbation.[27] The numerical difference between the two has been shown to be small.[6] Therefore also the short-range RPA + μ-SOSEX expressions obtained from the drCCD and ACFDT approaches are not identical. The numerical impact of this difference will be examined in Sec. IV.

III. COMPUTATIONAL DETAILS

We have implemented all RPA, RPA + SOSEX, and short-range RPA + μ-SOSEX routines in the FermiONs++ program package developed in our group.[35,36] For the drCCD type calculations, we have implemented two different approaches to evaluate the Riccati equation. The first method proposed by Heßelmann[28] uses an iterative, damped amplitude update scheme. Since we faced convergence problems with this method for small gap systems (e.g., the dissociation of molecules), we have also implemented a pseudo-Newton method[18] with a direct inversion of the iterative subspace (DIIS) scheme[37,38] adapted for CC calculations.[39] We used Heßelmann's approach where possible due to its simplicity and lower computational cost and resorted to the pseudo-Newton method for difficult cases. For Heßelmann's approach, we converged the correlation energies to 10^{-8} H, while for the pseudo-Newton method, we converged the norm of the residual of the Riccati equation also to 10^{-8} H.

For all calculations using the ACFDT formulation, we closely follow the implementation described in Ref. 10. This entails using a Clenshaw-Curtis quadrature for the numerical frequency quadrature adapted for RPA type calculations, originally proposed by Furche *et al.*[40] While it was shown that the adaption of the integration parameter developed originally for RPA[40] can be employed directly for closed shell RPA + SOSEX calculations,[10] this is not true for spin-unrestricted RPA + SOSEX or short-range RPA + SOSEX calculations. We found, however, that employing the same adaptive quadrature for unrestricted RPA + SOSEX and RPA + μ-SOSEX calculations also delivers improved convergence as compared to the conventional Clenshaw-Curtis quadrature.[41] We have used 256 integration points in most calculations resorting to 128 points for computationally demanding systems and to more integration points for difficult cases, in which the numerical integration was not converged with 256 points.

Both the drCCD and the ACFDT implementations used in this work scale as $N_{occ}^3 N_{virt}^3$, where N_{occ} and N_{virt} are the number of occupied and virtual orbitals, respectively. We note that there are a variety of approaches permitting more efficient RPA calculations.[28,40,42–47]

While there is an ongoing discussion about the best reference orbitals for RPA calculations (see, e.g., Ref. 48), we have performed all calculations using orbitals obtained from a DFT calculation employing the Perdew-Burke-Ernzerhof (PBE)[49] exchange-correlation functional, as is well established in the literature.[7,18,20,50] Total energies were obtained as usual by adding the correlation energy to the Hartree-Fock energy evaluated with the PBE reference orbitals (HXX energy).

We have assessed the performance of the short-range SOSEX correction on several test sets. Since one of the primary ideas was to overcome the problems associated with RPA + SOSEX barrier heights, we chose the BH76,[51,52] DBH24/08, HTBH38/08, and NHTBH38/08[53,54] test sets to thoroughly evaluate the performance for barrier heights. Furthermore, we performed calculations on the G21IP, G21EA and the W4-08 test sets[51,52] to see how the correction affects the description of non-isogyric processes, namely, ionisation potentials, electron affinities, and atomisation energies. We excluded reaction 44 from the W4-08 test set, due to problems in obtaining a reliable value for this reaction, as reported earlier in the literature for RPA calculations.[55] Finally, we also performed calculations on the SIE11[51,52] test set to evaluate the performance for systems, where molecular self-interaction errors play an important role.

In all our calculations, we used correlation consistent (cc-pVXZ) or augmented correlation consistent (aug-cc-pVXZ) basis sets.[56] In case complete basis set (CBS) extrapolation was applied, we used a procedure proposed earlier for RPA calculations[57] and then added the extrapolated correlation energies to the HXX energies, evaluated using the PBE reference orbitals of the higher cardinality basis to obtain total energies. We employed the frozen core approximation in all calculations.

IV. RESULTS AND DISCUSSION

A. H_2- and H_2^+-dissociation

As a first test for studying to which extent our short-range SOSEX correction enables the combination of removing the self-interaction present in RPA, while not sacrificing the good description of static correlation, we calculated H_2- and H_2^+-dissociation curves as paradigm cases for the impact of static correlation and self-interaction error.

As can be seen in Fig. 1, RPA underestimates the total energy at the equilibrium bond distance, while the dissociation limit is correctly described.[18,19] On the other hand, RPA + SOSEX drastically improves the description of short-range correlation energies, as can be seen by the overlap with the CCSD curve around the equilibrium bond distance. In the dissociation limit, however, RPA + SOSEX reintroduces the symmetry dilemma for stretched H_2 leading to a too high

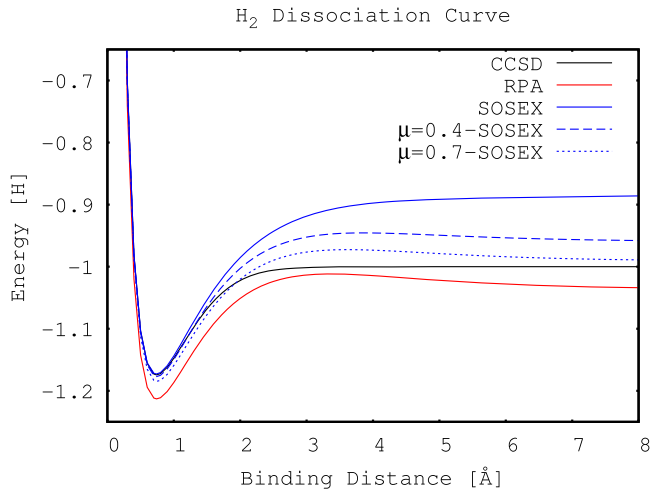


FIG. 1. Closed-shell H_2-dissociation curves for RPA, RPA + SOSEX (SOSEX), and RPA + μ-SOSEX (μ-SOSEX) calculations using a PBE reference and an aug-cc-pV5Z basis. The CCSD curve represents the exact curve. All RPA and beyond RPA curves were calculated using the drCCD approach.

dissociation energy. The proposed short-range SOSEX correction clearly reduces the large error of RPA + SOSEX in the long-range regime, while the good description around the equilibrium distance is retained. This is shown exemplarily in Fig. 1 for attenuation parameters $\mu = 0.4$ and $\mu = 0.7$.

The RPA + SOSEX dissociation curve for H_2^+ shown in Fig. 2 corresponds to the exact curve. The RPA curve deviates significantly from the exact curve at the equilibrium distance and in the dissociation limit due to the self-interaction error present in the correlation energy. The short-range SOSEX correction offers significant improvement over RPA; however, it also suffers from significant self-interaction in the dissociation limit, a phenomenon that is also present in alternative beyond-RPA approaches.[14] The degree to which the preceding conclusion applies is obviously again dependent on the attenuation parameter, where exemplarily $\mu = 0.4$ and $\mu = 0.7$ are shown in Fig. 2.

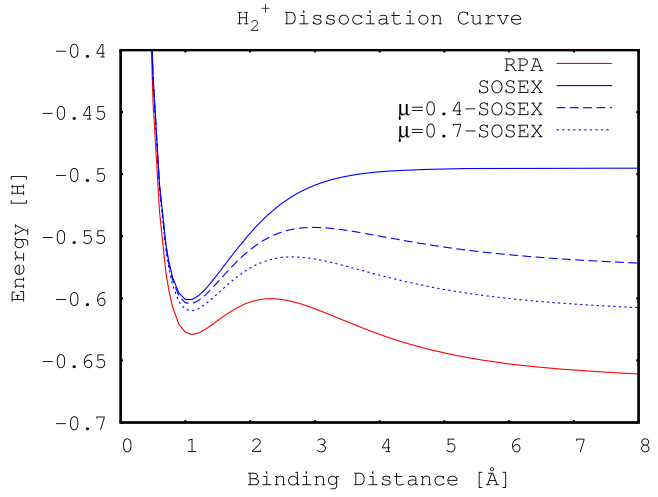


FIG. 2. H_2^+-dissociation curves for RPA, RPA + SOSEX (SOSEX), and RPA + μ-SOSEX (μ-SOSEX) calculations using a PBE reference and an aug-cc-pV5Z basis. The RPA + SOSEX curve represents the exact curve. All RPA and beyond RPA curves were calculated using the drCCD approach.

The results shown in this paragraph for the two paradigm systems for static correlation through bond dissociation and self-interaction error, namely, the dissociation of closed-shell H_2 and H_2^+, exemplify that the short-range SOSEX correction enables to interpolate between the complete removal of the one-electron self-interaction error, with the limiting case of RPA + SOSEX and the good description of static correlation, with the limiting case of plain RPA. Since both RPA and RPA + SOSEX have been shown to have drawbacks as reviewed in the introduction and the theory part, we hope to generate a post-Kohn-Sham functional that combines the benefits of both, leading to a balanced description along a wide range of properties. Therefore it would be necessary to perform a thorough parameterisation of the attenuation parameter μ. We do not perform this parameterisation in this work and set $\mu = 0.7$ from now on, to qualitatively show that this post-Kohn-Sham correction achieves good results for the properties examined in Secs. IV B–IV D, aside from the model systems studied so far.

B. Barrier heights and reaction energies

Next we performed calculations on barrier heights since RPA + SOSEX is known to deteriorate the good performance of plain RPA calculations and we aim to improve this shortcoming with our short-range RPA + μ-SOSEX approach. Therefore we chose the HTBH38/08 and NHTBH38/08 test sets[53,54] to evaluate the performance of the three methods for hydrogen transfer (HTBH38/08) and non-hydrogen transfer (NHTBH38/08) barrier heights. Both sets combined form the BH76 test set.[51,52] We used an aug-cc-pVQZ basis set for our calculations on the entire test set as recommended as a minimum size basis set for RPA calculations in the literature.[57] To make sure that our results are not qualitatively influenced by basis set incompleteness errors, we performed complete basis set (CBS) extrapolation of the correlation energies[57] for the DBH24/08 subset using aug-cc-pVQZ, aug-cc-pV5Z extrapolation. The DBH24/08 set forms a representative subset of the HTBH38/08 and NHTBH38/08 test sets. Barrier heights were calculated for both energies obtained from the drCCD and the ACFDT framework. The results are displayed in Table I. For the drCCD calculations, RPA + SOSEX shows the by far worst performance across all test sets. RPA shows remarkably good results throughout, as already mentioned earlier in the literature.[22] Our newly proposed short-range SOSEX correction, however, improves upon both RPA + SOSEX and plain RPA calculations for each set in Table I. When comparing the performance for hydrogen transfer vs. non-hydrogen transfer barrier heights, the results show that the short-range SOSEX correction performs especially well for non-hydrogen transfer barrier heights. The results for the DBH24/08 test set with the aug-cc-pVQZ, aug-cc-pV5Z basis sets show rather small differences to the CBS extrapolated results as compared to the differences between the three different methods. This gives confidence in the conclusions drawn from the results obtained with the smaller aug-cc-pVQZ basis.

As mentioned earlier, the RPA results calculated from the ACFDT approach are identical to the drCCD results, as has been proven analytically.[9] The fact that the mean absolute error (MAE) agrees to sub-0.01 kcal/mol accuracy gives

TABLE I. Mean absolute errors (MAEs) for barrier heights of the HTBH38/08, NHTBH38/08, DBH24/08, and BH76 test sets. The basis set used is indicated below the specification of the test set. The CBS results for the DBH24/08 test set were extrapolated using aug-cc-pVQZ, aug-cc-pV5Z basis set extrapolation. All results are in kcal/mol.

	HTBH38/08	NHTBH38/08	BH76	DBH24/08		
	aug-cc-pVQZ	aug-cc-pVQZ	aug-cc-pVQZ	aug-cc-pVQZ	aug-cc-pV5Z	CBS
drCCD						
RPA	1.71	1.96	1.83	1.68	1.69	1.70
RPA + SOSEX	5.16	4.13	4.64	4.04	3.92	3.81
RPA + μ = 0.7-SOSEX	1.61	1.22	1.41	1.24	1.15	1.22
ACFDT						
RPA	1.71	1.96	1.83	1.68	1.69	1.70
RPA + SOSEX	5.47	4.37	4.92	4.27	4.16	4.05
RPA + μ = 0.7-SOSEX	1.69	1.24	1.46	1.27	1.17	1.24

confidence in our implementation of the different methods since it enables to cross-validate the amplitude and quadrature based calculations. The ACFDT based RPA + SOSEX energy expressions on the other hand are not identical.[27] Previous tests have shown, however, that the numerical difference is small.[6] This is confirmed by the results presented here. The MAEs obtained from ACFDT calculations qualitatively agree with the drCCD results, with a maximum deviation of 0.3 kcal/mol in the MAE. The same holds true for the proposed RPA + μ-SOSEX, with a maximum deviation of 0.1 kcal/mol. In terms of individual results, all properties calculated in this section and Secs. IV C and IV D agree with a maximum difference of approximately 1 kcal/mol between the ACFDT and drCCD results.

Having calculated the forward and backward barrier heights of the HTBH and NHTBH test sets, together constituting the BH76 test set, one can calculate 30 reaction energies, which are summarised in the BH76RC test set.[51,52] At the aug-cc-pVQZ level of theory, RPA provides the worst results for reaction energies with an MAE of 2.3. In contrast to barrier heights, the results for reaction energies of RPA + SOSEX are slightly better than the plain RPA calculations with an MAE of 2.1 kcal/mol for the drCCD calculations (ACFDT result: 2.2 kcal/mol). This can possibly be accounted for the fact that the reaction reagents and products do not exhibit strong static correlation effects. However also for reaction energies, our newly proposed RPA + μ-SOSEX provides the best results with an MAE of 1.6 kcal/mol for both the ACFDT and drCCD calculations.

C. Non-isogyric processes: Ionisation potentials, electron affinities, and atomisation energies

In the previous paragraph, we showed that the attenuated SOSEX correction improves upon both the RPA and conventional RPA + SOSEX results. While for barrier heights, RPA provides significantly better results than RPA + SOSEX, the opposite holds true for non-isogyric processes. To evaluate the performance of our proposed attenuated SOSEX correction for electron affinities and ionisation potentials, we performed calculations on the G21EA and G21IP test sets.[51,52] In order to obtain results free from basis set incompleteness errors, we performed an aug-cc-pVQZ, aug-cc-pV5Z extrapolation for the G21EA test set and a cc-pVQZ, cc-pV5Z extrapolation for the G21IP test set. The reason why we moved to non-augmented basis sets for the ionisation potential test set is that there are no aug-cc-pV5Z basis sets available for Li, Na, Mg, and Be. The results presented in Table II show that the RPA results for both ionisation potentials and electron affinities are the worst, while RPA + SOSEX delivers the best results. Our short-range RPA + μ-SOSEX correction is significantly better than plain RPA, however slightly worse than the conventional RPA + SOSEX results. In particular for the drCCD formalism, the MAE of our short-range RPA + μ-SOSEX results is approximately 4 kcal/mol better than RPA and approximately 1 kcal/mol worse than RPA + SOSEX for electron affinities in the complete basis set limit. For ionisation potentials, the MAE of the short-range RPA + μ-SOSEX correction is approximately 5 kcal/mol better than RPA and again approximately 1 kcal/mol worse than RPA + SOSEX. The results for pentuple zeta basis sets show that both RPA and our short-range RPA + μ-SOSEX correction seem to benefit from an error compensation of methodological error and basis set incompleteness error. The results from the

TABLE II. MAEs for electron affinities and ionisation potentials of the G21EA and G21IP test sets. The CBS extrapolation for electron affinities was performed using aug-cc-pVQZ, aug-cc-pV5Z basis sets, while the results for ionisation potentials were obtained using cc-pVQZ, cc-pV5Z basis set extrapolation. All results are in kcal/mol.

	G21EA		G21IP	
	aug-cc-pV5Z	CBS	cc-pV5Z	CBS
drCCD				
RPA	6.26	7.31	6.48	8.07
RPA + SOSEX	3.29	2.79	2.55	2.65
RPA + μ = 0.7-SOSEX	3.01	3.67	2.81	3.56
ACFDT				
RPA	6.26	7.31	6.48	8.07
RPA + SOSEX	3.42	2.93	2.62	2.41
RPA + μ = 0.7-SOSEX	3.02	3.69	2.83	3.57

ACFDT algorithms for RPA calculations again agree to below 0.01 kcal/mol, while the RPA + SOSEX and short-range RPA + SOSEX results show deviations of up to 0.2 kcal/mol but qualitatively lead to the same conclusions. To summarize, the results of this paragraph show that for ionisation potentials and electron affinities, the full RPA + SOSEX correction performs slightly better than our short-range SOSEX correction, which in turn performs significantly better than plain RPA.

As a last test for non-isogyric processes, we examined the performance of RPA + μ-SOSEX for the W4-08 atomisation test set[51,52] and the W4-08MR subset, consisting of systems which exhibit a strong degree of multi-reference character. It is well known that for atomization energies, RPA consistently underbinds.[7,58,59] It was shown that RPA + SOSEX corrects this issue to some extent.[58,59] The same can be seen here for the W4-08 test set as shown by the results presented in Table III. RPA shows the worst performance for atomization energies, while RPA + SOSEX shows the best results. The short-range RPA + μ-SOSEX correction performs better than RPA but worse than RPA + SOSEX. While for ionisation potentials and electron affinities, the difference between the short-range RPA + μ-SOSEX and conventional RPA + SOSEX was small in comparison to the performance of plain RPA, this is not the case for the entire W4-08 test set. This indicates that for atomisation energies, a full-range SOSEX correction seems to be beneficial. Interestingly, however, for the subset of multi-reference cases, plain RPA shows the best results, while RPA + μ-SOSEX performs slightly worse. Both of them exhibit significantly better results than RPA + SOSEX. This is in line with the observation that RPA describes static correlation better than RPA + SOSEX and confirms that our short-range RPA + μ-SOSEX combines the description of static correlation and removal of self-interaction error. On an absolute scale, all results for atomisation energies are somehow unsatisfactory. There are, however, orthogonal correction schemes that improve this shortcoming such as the renormalized singles correction, which have been shown to deliver improved atomization energies.[22]

TABLE III. MAEs for atomization energies of the W4-08 and the W4-08MR test set. All results were obtained with CBS extrapolated correlation energies. The CBS extrapolation of the correlation energies was performed using cc-pVQZ, cc-pV5Z basis sets. All results are in kcal/mol.

	drCCD		ACFDT	
	W4-08	W4-08MR	W4-08	W4-08MR
RPA	12.43	9.60	12.43	9.60
RPA + SOSEX	9.06	19.33	9.09	19.66
RPA + μ = 0.7-SOSEX	11.21	12.46	11.07	12.35

D. Self-interaction error

As a final test for our short-range RPA + μ-SOSEX correction, we evaluated and compared its performance on the SIE11 test set.[51,52] This set contains reactions that are prone to self-interaction errors. We have already examined the performance of the three methods for the one-electron self-interaction error using the example of stretched H_2^+ in Sec. IV A. While RPA + SOSEX is self-interaction free for one-electron systems, this is not true for more complicated systems. Therefore we benchmark the performance of the three methods for molecular systems in this section. Because the last paragraphs have shown that the results obtained from the drCCD calculations are very similar to the results obtained from ACFDT calculations, we have performed calculations of the SIE11 test set only with the drCCD variants, due to the significantly decreased computational cost of solving the Riccati equation as compared to the numerical frequency integration of the ACFDT based methods. The results presented in Table IV show that RPA + SOSEX does not improve the MAE for the SIE11 test set significantly as compared to a plain RPA calculation. Similar results have been reported earlier in the literature.[60] RPA

TABLE IV. Deviations from the reference results of drCCD RPA, RPA + SOSEX, and RPA + μ-SOSEX calculations for the eleven reactions contained in the SIE11 test set. All energies were extrapolated to the CBS limit using cc-pVQZ, cc-pV5Z extrapolation. Additionally mean errors (MEs), mean absolute errors (MAEs), and minimum (Min) and maximum (Max) errors are given for each method for the entire test set. All results are in kcal/mol.

Reagents	Products	RPA	RPA + SOSEX	RPA + μ = 0.7-SOSEX
He_2^+	He + He^+	13.43	−5.64	0.29
$(NH_3)_2^+$	NH_3 + NH_3^+	4.45	−5.85	−0.38
$(H_2O)_2^+$	H_2O + H_2O^+	9.49	−8.27	1.34
$(C_4H_{10})^+$	C_2H_5 + $C_2H_5^+$	2.44	−4.65	−1.44
$(CH_3)_2CO^+$	CH_3 + CH_3CO^+	0.86	−0.01	−0.32
ClFCl	ClClF	7.61	−3.42	4.66
$C_2H_4\cdots F_2$	C_2H_4 + F_2	−2.69	−3.04	−2.66
$C_6H_6\cdots Li$	C_6H_6 + Li	−8.94	−7.74	−12.87
$NH_3\cdots ClF$	NH_3 + ClF	−0.88	−2.42	−1.70
NaOMg	MgO + Na	−0.15	11.50	0.41
FLiF	F_2 + Li	44.27	−42.68	−0.10
	MAE	8.66	8.66	2.38
	ME	6.35	−6.57	−1.16
	Min	−8.94	−42.68	−12.87
	Max	44.27	11.50	4.66

consistently overbinds, while inclusion of the SOSEX correction leads to systematic underbinding. On the other hand, our short-range RPA + SOSEX correction shows a mean absolute error of only 2.4 kcal/mol. The negative mean error indicates a tendency to underbind, however, not to the same extent as RPA + SOSEX. This shows that while RPA + SOSEX is self-interaction free for one-electron systems, its performance for molecular systems prone to self-interaction errors is not better than plain RPA, which suffers from serious self-correlation. Our short-range SOSEX correction on the other hand offers an improvement over RPA for the one-electron case (see Sec. IV A) and a significant improvement over RPA and RPA + SOSEX for the molecular systems' self-interaction benchmark shown in this paragraph.

V. CONCLUSION

We have introduced a scheme that provides a well balanced compromise between plain RPA and RPA + SOSEX for the calculation of the correlation energy. This enables to combine the benefits of both methods, namely, removal of self-interaction errors and the description of static correlation. We have shown this for the two paradigm cases of self-interaction errors and static correlation, i.e., the dissociation of H_2^+ and closed shell H_2. For real world systems, we have shown that our short-range RPA + SOSEX correction, without extensive parameterisation, improves upon both RPA and RPA + SOSEX barrier heights and reaction energies, while ionisation potentials and electron affinities are slightly worse than RPA + SOSEX but significantly better than the RPA results. Only for atomisation energies, the full RPA + SOSEX correction seems to be necessary for significant improvement over RPA. One exception here is multi-reference cases, where both RPA and our short-range RPA + μ-SOSEX deliver significantly improved results over RPA + SOSEX. Finally we showed that for molecular systems prone to self-interaction errors, our short-range SOSEX correction clearly outperforms both RPA and RPA + SOSEX. The entirety of our results confirms that a short-range RPA + SOSEX correction enables to combine the benefits of RPA and RPA + SOSEX calculations, delivering a balanced description across a variety of properties.

Future work on this subject should start with a thorough parameterisation of the attenuation parameter μ. Further performance increases might be possible by combining this approach with orthogonal correction schemes, such as the renormalized singles excitation correction.[22]

When comparing the drCCD based variant with the ACFDT version, the former seems superior in terms of accuracy and speed when considering the molecular orbital implementations since no matrix diagonalisations are necessary and the number of iterations necessary for solving the Riccati equation is usually smaller than the number of frequency points necessary for the numerical quadrature. However, when switching to an atomic-orbital or a Cholesky basis, the ACFDT version of RPA has shown potential for great computational savings.[42,43] Therefore an ACFDT based atomic-orbital variant of the short-range RPA + SOSEX correction might be a viable approach for calculations of large molecules. Here, the inherent locality of the short-range SOSEX correction could also be beneficial.

ACKNOWLEDGMENTS

The authors thank Dr. A. Luenser and Dr. H. F. Schurkus (LMU Munich) for helpful discussions. Financial support was provided by the Excellence Cluster EXC114 (CIPSM) and SFB749 by the Deutsche Forschungsgemeinschaft (DFG).

[1] X. Ren, P. Rinke, C. Joas, and M. Scheffler, J. Mater. Sci. **47**, 7447 (2012).
[2] H. Eshuis, J. E. Bates, and F. Furche, Theor. Chem. Acc. **131**, 1084 (2012).
[3] G. P. Chen, V. K. Voora, M. M. Agee, S. G. Balasubramani, and F. Furche, Annu. Rev. Phys. Chem. **68**, 421 (2017).
[4] D. Langreth and J. Perdew, Solid State Commun. **17**, 1425 (1975).
[5] D. C. Langreth and J. P. Perdew, Phys. Rev. B **15**, 2884 (1977).
[6] J. G. Ángyán, R.-F. Liu, J. Toulouse, and G. Jansen, J. Chem. Theory Comput. **7**, 3116 (2011).
[7] F. Furche, Phys. Rev. B **64**, 195120 (2001).
[8] F. Furche, J. Chem. Phys. **129**, 114105 (2008).
[9] G. E. Scuseria, T. M. Henderson, and D. C. Sorensen, J. Chem. Phys. **129**, 231101 (2008).
[10] B. Mussard, D. Rocca, G. Jansen, and J. G. Ángyán, J. Chem. Theory Comput. **12**, 2191 (2016).
[11] A. Heßelmann and A. Görling, Mol. Phys. **108**, 359 (2010).
[12] A. Szabo and N. S. Ostlund, J. Chem. Phys. **67**, 4351 (1977).
[13] A. Szabo and N. S. Ostlund, Int. J. Quantum Chem. **12**, 389 (1977).
[14] J. E. Bates and F. Furche, J. Chem. Phys. **139**, 171103 (2013).
[15] A. D. McLachlan and M. A. Ball, Rev. Mod. Phys. **36**, 844 (1964).
[16] J. Toulouse, I. C. Gerber, G. Jansen, A. Savin, and J. G. Ángyán, Phys. Rev. Lett. **102**, 096404 (2009).
[17] J. Toulouse, W. Zhu, J. G. Ángyán, and A. Savin, Phys. Rev. A **82**, 032502 (2010).
[18] T. M. Henderson and G. E. Scuseria, Mol. Phys. **108**, 2511 (2010).
[19] M. Fuchs, Y.-M. Niquet, X. Gonze, and K. Burke, J. Chem. Phys. **122**, 094116 (2005).
[20] J. F. Dobson, J. Wang, B. P. Dinte, K. McLennan, and H. M. Le, Int. J. Quantum Chem. **101**, 579 (2005).
[21] J. F. Dobson and T. Gould, J. Phys.: Condens. Matter **24**, 073201 (2012).
[22] X. Ren, P. Rinke, G. E. Scuseria, and M. Scheffler, Phys. Rev. B **88**, 035120 (2013).
[23] Z. Yan, J. P. Perdew, and S. Kurth, Phys. Rev. B **61**, 16430 (2000).
[24] J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse, and M. Scheffler, New J. Phys. **14**, 043002 (2012).
[25] A. Grüneis, M. Marsman, J. Harl, L. Schimka, and G. Kresse, J. Chem. Phys. **131**, 154115 (2009).
[26] D. L. Freeman, Phys. Rev. B **15**, 5512 (1977).
[27] G. Jansen, R.-F. Liu, and J. G. Ángyán, J. Chem. Phys. **133**, 154106 (2010).
[28] A. Heßelmann, Phys. Rev. A **85**, 012517 (2012).
[29] W. Zhu, J. Toulouse, A. Savin, and J. G. Ángyán, J. Chem. Phys. **132**, 244108 (2010).
[30] B. G. Janesko, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **130**, 081105 (2009).
[31] B. G. Janesko, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **131**, 034110 (2009).
[32] R. M. Irelan, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **135**, 094105 (2011).
[33] B. Mussard, P. Reinhardt, J. G. Ángyán, and J. Toulouse, J. Chem. Phys. **142**, 154123 (2015).
[34] J. J. Shepherd, T. M. Henderson, and G. E. Scuseria, Phys. Rev. Lett. **112**, 133002 (2014).
[35] J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013).
[36] J. Kussmann and C. Ochsenfeld, J. Chem. Theory Comput. **11**, 918 (2015).
[37] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[38] P. Pulay, J. Comput. Chem. **3**, 556 (1982).
[39] G. E. Scuseria, T. J. Lee, and H. F. Schaefer, Chem. Phys. Lett. **130**, 236 (1986).
[40] H. Eshuis, J. Yarkony, and F. Furche, J. Chem. Phys. **132**, 234114 (2010).
[41] J. P. Boyd, J. Sci. Comput. **2**, 99 (1987).
[42] H. F. Schurkus and C. Ochsenfeld, J. Chem. Phys. **144**, 031101 (2016).
[43] A. Luenser, H. F. Schurkus, and C. Ochsenfeld, J. Chem. Theory Comput. **13**, 1647 (2017).

[44]M. Kállay, J. Chem. Phys. **142**, 204105 (2015).
[45]A. Heßelmann, J. Chem. Phys. **146**, 174110 (2017).
[46]J. Wilhelm, P. Seewald, M. Del Ben, and J. Hutter, J. Chem. Theory Comput. **12**, 5851 (2016).
[47]M. Kaltak, J. Klimeš, and G. Kresse, J. Chem. Theory Comput. **10**, 2498 (2014).
[48]J. E. Bates, P. D. Mezei, G. I. Csonka, J. Sun, and A. Ruzsinszky, J. Chem. Theory Comput. **13**, 100 (2017).
[49]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[50]J. Harl and G. Kresse, Phys. Rev. B **77**, 045136 (2008).
[51]L. Goerigk and S. Grimme, J. Chem. Theory Comput. **6**, 107 (2010).
[52]L. Goerigk and S. Grimme, Phys. Chem. Chem. Phys. **13**, 6670 (2011).
[53]J. Zheng, Y. Zhao, and D. G. Truhlar, J. Chem. Theory Comput. **5**, 808 (2009).
[54]R. Peverati and D. G. Truhlar, Philos. Trans. R. Soc., A **372**, 20120476 (2014).
[55]S. Grimme and M. Steinmetz, Phys. Chem. Chem. Phys. **18**, 20926 (2016).
[56]T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[57]H. Eshuis and F. Furche, J. Chem. Phys. **136**, 084105 (2012).
[58]J. Paier, B. G. Janesko, T. M. Henderson, G. E. Scuseria, A. Grüneis, and G. Kresse, J. Chem. Phys. **132**, 094103 (2010).
[59]J. Paier, B. G. Janesko, T. M. Henderson, G. E. Scuseria, A. Grüneis, and G. Kresse, J. Chem. Phys. **133**, 179902 (2010).
[60]A. Ruzsinszky, I. Y. Zhang, and M. Scheffler, J. Chem. Phys. **143**, 144115 (2015).

3.5 Publication V: Efficient calculation of beyond RPA correlation energies in the dielectric matrix formalism

M. Beuerle, D. Graf, H. F. Schurkus, C. Ochsenfeld,
"Efficient Calculation of Beyond RPA Correlation Energies in the Dielectric Matrix Formalism",
J. Chem. Phys., **148**, 204104 (2018).

Abstract: We present efficient methods to calculate beyond random phase approximation (RPA) correlation energies for molecular systems with up to 500 atoms. To reduce the computational cost, we employ the resolution-of-the-identity and a double-Laplace transform of the non-interacting polarization propagator in conjunction with an atomic orbital formalism. Further improvements are achieved using integral screening and the introduction of Cholesky decomposed densities. Our methods are applicable to the dielectric matrix formalism of RPA including second-order screened exchange (RPA-SOSEX), the RPA electron-hole time-dependent Hartree-Fock (RPA-eh-TDHF) approximation, and RPA renormalized perturbation theory using an approximate exchange kernel (RPA-AXK). We give an application of our methodology by presenting RPA-SOSEX benchmark results for the L7 test set of large, dispersion dominated molecules, yielding a mean absolute error below 1 kcal/mol. The present work enables calculating beyond RPA correlation energies for significantly larger molecules than possible to date, thereby extending the applicability of these methods to a wider range of chemical systems.

THE JOURNAL OF CHEMICAL PHYSICS **148**, 204104 (2018)

Efficient calculation of beyond RPA correlation energies in the dielectric matrix formalism

Matthias Beuerle, Daniel Graf, Henry F. Schurkus, and Christian Ochsenfeld[a)]
Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 München, Germany and Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5–13, D-81377 München, Germany

(Received 14 February 2018; accepted 25 April 2018; published online 23 May 2018)

We present efficient methods to calculate beyond random phase approximation (RPA) correlation energies for molecular systems with up to 500 atoms. To reduce the computational cost, we employ the resolution-of-the-identity and a double-Laplace transform of the non-interacting polarization propagator in conjunction with an atomic orbital formalism. Further improvements are achieved using integral screening and the introduction of Cholesky decomposed densities. Our methods are applicable to the dielectric matrix formalism of RPA including second-order screened exchange (RPA-SOSEX), the RPA electron-hole time-dependent Hartree-Fock (RPA-eh-TDHF) approximation, and RPA renormalized perturbation theory using an approximate exchange kernel (RPA-AXK). We give an application of our methodology by presenting RPA-SOSEX benchmark results for the L7 test set of large, dispersion dominated molecules, yielding a mean absolute error below 1 kcal/mol. The present work enables calculating beyond RPA correlation energies for significantly larger molecules than possible to date, thereby extending the applicability of these methods to a wider range of chemical systems. *Published by AIP Publishing.* https://doi.org/10.1063/1.5025938

I. INTRODUCTION

Correlation energies obtained from the direct random phase approximation (dRPA) have proven to be a valuable post-Kohn-Sham (KS) correction (for an overview over recent developments, see, e.g., Refs. 1–3). While the original formulations for calculating dRPA correlation energies for molecules showed an $\mathcal{O}(\mathcal{N}^6)$ asymptotic scaling behavior,[4] where N denotes the molecular size, restricting the application to small molecules, several reformulations have been introduced recently for reducing the scaling behavior and allowing for calculating larger systems, some with more than 1000 atoms.[5–8] To reduce the computational cost, these methods employ, e.g., the resolution-of-the-identity (RI) technique,[9] tensor hyper contraction (THC),[10–13] local,[5] atomic,[6] and Cholesky orbital[7] formulations, and integral transforms of the non-interacting polarization propagator,[6–8,14,15] the latter being a central quantity in the calculation of dRPA correlation energies. For some methods, this brought the asymptotic scaling behavior down to linear,[5–8] enabling calculations for large systems of chemical interest.

While dRPA calculations provide significant improvements over conventional density functional theory (DFT) calculations for properties such as dispersion interactions,[16,17] dRPA delivers unsatisfactory results for non-isogyric processes such as atomization energies.[4,18–20] These failures can be traced back to the self-interaction error present in dRPA energies.[21] To circumvent these problems, approaches beyond the direct random phase approximation have been proposed that include exchange effects and higher order correlations.

In general, RPA-type methods can be derived from two frameworks, namely, those that resemble simplified coupled cluster doubles equations[22] or those derived from the adiabatic connection fluctuation dissipation theorem (ACFDT).[4,23–25] Here we focus on approaches derived from the latter framework. In the ACFDT, the direct random phase approximation represents the simplest approach to obtain an approximate interacting polarization propagator required for the calculation of ACFDT correlation energies. To further include exchange effects, several approaches have been suggested: Among those, a second-order screened exchange (RPA-SOSEX) type approach[26,27] replaces the Hartree kernel in the ACFDT formula with an antisymmetrised Hartree kernel.[28,29] The RPA electron-hole time-dependent Hartree-Fock (RPA-eh-TDHF) approximation uses an interacting polarization propagator obtained from a simplified time-dependent Hartree-Fock kernel.[29,30] RPA renormalized perturbation theory uses a low-order approximation to the approximate eh-TDHF polarization propagator as the leading correction to dRPA (this approach is denoted as RPA-AXK).[18] Furthermore, methods have been proposed using the exact-exchange kernel from time-dependent density functional theory,[31] also including a power series approximation to the correlation kernel.[32] Applications of these methods have been hampered so far by their steep computational scaling of up to $\mathcal{O}(\mathcal{N}^6)$ in their canonical formulation. Using the RI-technique, $\mathcal{O}(\mathcal{N}^5)$-scaling ACFDT-based formulations have been proposed,[29,30,33] which is however still too expensive to tackle

[a)]christian.ochsenfeld@uni-muenchen.de

large molecular systems. Further performance benefits have been obtained using a plane wave formulation in conjunction with a Gram-Schmidt orthogonalization scheme.[34] It is worthwhile to note that more efficient RPA with exchange methods have been proposed not only in the ACFDT but also the CCD framework.[35,36]

Here we present a framework that allows for a low scaling calculation of different RPA with exchange methods. In particular, we show that our methods are applicable for RPA-SOSEX, RPA-eh-TDHF, and RPA-AXK. We employ methods recently introduced by us in the context of low- and linear-scaling dRPA and MP2 methods,[6–8,37] namely, an RI-decomposition using a local metric in conjunction with an atomic orbital (AO) or Cholesky basis formulation using the integrated double-Laplace transform of the non-interacting polarization propagator,[6] which is equivalent to a Fourier transform of the non-interacting polarization propagator into the imaginary frequency domain.[14] In this way, RPA with exchange energies for significantly larger molecular systems become accessible.

II. THEORY

In the following, we present a derivation of the working equations for this manuscript. For detailed information on the different RPA with exchange methods and their derivations, we refer the reader to the original publications.[18,29,30] We use the Mulliken notation for two, three, and four center Coulomb integrals and enumerate quantities related to occupied orbitals as i, j, ... and those related to virtual orbitals as a, b, ...

The dRPA correlation energy in the ACFDT using real-valued spin orbitals after coupling strength integration is given as[23,24,29]

$$E_c^{\mathrm{dRPA}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)+\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right\}, \quad (1)$$

with $\mathbf{\Pi}_0(\mathrm{i}\omega)$ being the non-interacting polarization propagator at imaginary frequency $\mathrm{i}\omega$,

$$\Pi_0(\mathrm{i}\omega)_{ia,jb} = \frac{-2\varepsilon_{ia}}{\varepsilon_{ia}^2+\omega^2}\delta_{i,j}\delta_{a,b}, \quad (2)$$

and $V_{ia,jb} = (ia|jb)$ being the Hartree kernel. Here $\varepsilon_{ia} = \varepsilon_a - \varepsilon_i$ are particle-hole excitation energies, where ε_i, ε_j, ... denote the occupied orbital energies and ε_a, ε_b, ... denote the virtual orbital energies. Turning to the three beyond RPA methods used in this work, the canonical formulation for the SOSEX correction is given by[29]

$$E_c^{\mathrm{SOSEX}} = -\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)\mathbf{V}^{-1}\mathbf{K} + \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\right\}, \quad (3)$$

the RPA-AXK correction is given by[18]

$$E_c^{\mathrm{AXK}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)\mathbf{V}^{-1}\mathbf{K} + \left(\mathbf{1}-\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)^{-1}\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\right\}, \quad (4)$$

and the correlation energy expression for RPA-eh-TDHF is given by[29,30]

$$E_c^{\text{RPA-eh-TDHF}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{W}\right)\mathbf{W}^{-1}\mathbf{V} + \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right\}. \quad (5)$$

Here $K_{ia,jb} = (ib|ja)$ represents an approximate exchange kernel and $\mathbf{W} = \mathbf{V} - \mathbf{K}$. It is worth to note that the SOSEX and AXK corrections have to be added to the dRPA correlation energy, while the formula for RPA-eh-TDHF delivers the entire correlation energy.

A. Resolution-of-the-identity formulations

The resolution-of-the-identity approximation is often used in quantum chemistry to decompose the four-index two-electron repulsion integral (ERI) tensor into two three-center tensors,[7,9,38–40]

$$(ia|jb) \approx \sum_{P,Q}(ia|m_{12}|P)C_{PQ}(Q|m_{12}|jb) = \sum_{P,Q,R}(ia|m_{12}|P)C_{PR}^{1/2}C_{RQ}^{1/2}(Q|m_{12}|jb) = \sum_R B_{ia}^R B_{jb}^R, \quad (6)$$

where $(ia|m_{12}|P)$ are the three-center integrals and the RI matrix

$$C_{PQ} = \sum_{R,S}(P|m_{12}|R)^{-1}(R|S)(S|m_{12}|Q)^{-1}, \quad (7)$$

with the respective metric m_{12} employed for the RI-decomposition. Here P, Q, R, ... denote the auxiliary basis functions.

Inserting the RI-approximation to decompose $\mathbf{V}$ leads to the following RI-dRPA expression introduced by Furche and co-workers:[9]

$$E_c^{\text{RI-dRPA}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{Q}(\mathrm{i}\omega)\right)+\mathbf{Q}(\mathrm{i}\omega)\right\}, \quad (8)$$

where $\mathbf{Q}(\mathrm{i}\omega)$ is a $N_{\mathrm{Aux}} \times N_{\mathrm{Aux}}$ matrix, with N_{Aux} being the size of the auxiliary space. $\mathbf{Q}(\mathrm{i}\omega)$ is defined as

$$Q_{PQ}(\mathrm{i}\omega) = \sum_{i,a} B_{ia}^P \Pi_0(\mathrm{i}\omega)_{ia,ia} B_{ia}^Q. \quad (9)$$

Similar considerations allow decomposing both $\mathbf{V}$ and $\mathbf{K}$ with the RI-approximation for the beyond RPA expressions to obtain

$$E_c^{\text{RI-SOSEX}} = -\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{Q}(\mathrm{i}\omega)\right)\mathbf{Q}^{-1}(\mathrm{i}\omega) \times \mathbf{Y}(\mathrm{i}\omega)\mathbf{Q}^{-1}(\mathrm{i}\omega) + \mathbf{Y}(\mathrm{i}\omega)\mathbf{Q}^{-1}(\mathrm{i}\omega)\right\}, \quad (10)$$

$$E_c^{\text{RI-AXK}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{Q}(\mathrm{i}\omega)\right)\mathbf{Q}^{-1}(\mathrm{i}\omega)\mathbf{Y}(\mathrm{i}\omega) \times \mathbf{Q}^{-1}(\mathrm{i}\omega) + \left(\mathbf{1}-\mathbf{Q}(\mathrm{i}\omega)\right)^{-1}\mathbf{Y}(\mathrm{i}\omega)\mathbf{Q}^{-1}(\mathrm{i}\omega)\right\}, \quad (11)$$

$$E_c^{\text{RI-RPA-eh-TDHF}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\left\{\log\left(\mathbf{1}-\mathbf{Q}(\mathrm{i}\omega)+\mathbf{Y}(\mathrm{i}\omega)\mathbf{Q}^{-1}(\mathrm{i}\omega)\right) \times \left(\mathbf{Q}(\mathrm{i}\omega)-\mathbf{Y}(\mathrm{i}\omega)\mathbf{Q}^{-1}(\mathrm{i}\omega)\right)^{-1}\mathbf{Q}(\mathrm{i}\omega)+\mathbf{Q}(\mathrm{i}\omega)\right\}. \quad (12)$$

The $\mathbf{Q}(\mathrm{i}\omega)$ matrix is the same as the one appearing in the dRPA energy expression [Eq. (8)] and $\mathbf{Y}(\mathrm{i}\omega)$ is given as

$$Y_{PQ}(\mathrm{i}\omega) = \sum_{i,a,j,b} B^P_{ia}\Pi_0(\mathrm{i}\omega)_{ia,ia}K_{ia,jb}\Pi_0(\mathrm{i}\omega)_{jb,jb}B^Q_{jb}. \quad (13)$$

For the derivation of the RI-expression for RPA-SOSEX, see Ref. 29; for RPA-eh-TDHF, see Ref. 30; and for RPA-AXK,[18] see the Appendix.

B. Atomic orbital formulation

While in Refs. 6–8, efficient methods to calculate $\mathbf{Q}(\mathrm{i}\omega)$ have been presented, the naive calculation of $\mathbf{Y}(\mathrm{i}\omega)$ scales at least as $\mathcal{O}(\mathcal{N}^2_{\mathrm{occ}}\mathcal{N}^2_{\mathrm{virt}}\mathcal{N}_{\mathrm{Aux}})$, where N_{occ} and N_{virt} denote the dimensions of the occupied and virtual space, respectively. Although this is more efficient than the canonical $\mathcal{O}(\mathcal{N}^3_{\mathrm{occ}}\mathcal{N}^3_{\mathrm{virt}})$ formulation for the beyond RPA methods, this scaling is still prohibitively steep for large molecules. To obtain an efficient method for calculating $\mathbf{Y}(\mathrm{i}\omega)$, we use a double-Laplace transform of the non-interacting polarization propagator,[6,14]

$$\begin{aligned}\Pi_0(\mathrm{i}\omega)_{ia,ia} &= \frac{-2\varepsilon_{ia}}{\varepsilon^2_{ia}+\omega^2} = -2\int_0^\infty \mathrm{d}\tau \cos(\omega\tau)e^{-\varepsilon_{ia}\tau} \\ &= -2\sum_\tau w_\tau \cos(\omega\tau)e^{-\varepsilon_{ia}\tau}, \end{aligned} \quad (14)$$

with roots τ and weights w_τ, which allows calculating $\mathbf{Y}(\mathrm{i}\omega)$ in an atomic orbital (AO) formulation analogous to the AO formulation of $\mathbf{Q}(\mathrm{i}\omega)$ first shown in Ref. 6,

$$\begin{aligned}Y_{PQ}(\mathrm{i}\omega) = &\sum_{\mu,\nu,\lambda,\sigma}\sum_{\mu',\nu',\lambda',\sigma'}\sum_{\tau,\tau'} 4w_\tau w_{\tau'}\cos(\omega\tau)\cos(\omega\tau') \\ &\times \underline{P}^\tau_{\mu\mu'}B^P_{\mu\nu}\overline{P}^\tau_{\nu\nu'}(\mu'\sigma'|\lambda'\nu')\underline{P}^{\tau'}_{\lambda\lambda'}B^Q_{\lambda\sigma}\overline{P}^{\tau'}_{\sigma\sigma'}, \end{aligned} \quad (15)$$

with $\underline{\mathbf{P}}^\tau$, $\overline{\mathbf{P}}^\tau$ representing the occupied and virtual pseudodensities, defined as

$$\underline{P}^\tau_{\mu\mu'} = \sum_i \mathcal{C}_{\mu i}e^{\varepsilon_i\tau}\mathcal{C}_{\mu' i} \quad (16)$$

and

$$\overline{P}^\tau_{\mu\mu'} = \sum_a \mathcal{C}_{\mu a}e^{-\varepsilon_a\tau}\mathcal{C}_{\mu' a}, \quad (17)$$

where $\mathcal{C}$ is the matrix of molecular orbital coefficients. We denote the AO basis functions as μ, μ', ν, ν', ...

The key idea of AO-based methods is a reformulation in a local Gaussian basis to obtain sparse quantities, which allow for an efficient calculation. Therefore, we also move the RI matrix $\mathbf{C}$ out of the $\mathbf{B}$ tensor[6] to obtain

$$Y_{PQ}(\mathrm{i}\omega) = \sum_{R,S} C^{1/2}_{PR}\tilde{Y}_{RS}(\mathrm{i}\omega)C^{1/2}_{SQ}, \quad (18)$$

$$\begin{aligned}\tilde{Y}_{PQ}(\mathrm{i}\omega) = &\sum_{\mu,\nu,\lambda,\sigma}\sum_{\mu',\nu',\lambda',\sigma'}\sum_{\tau,\tau'} 4w_\tau w_{\tau'}\cos(\omega\tau)\cos(\omega\tau') \\ &\times \underline{P}^\tau_{\mu\mu'}\tilde{B}^P_{\mu\nu}\overline{P}^\tau_{\nu\nu'}(\mu'\sigma'|\lambda'\nu')\underline{P}^{\tau'}_{\lambda\lambda'}\tilde{B}^Q_{\lambda\sigma}\overline{P}^{\tau'}_{\sigma\sigma'}, \end{aligned} \quad (19)$$

where $\tilde{B}^P_{\mu\nu} = (\mu\nu|m_{12}|P)$ represent solely the three-center integrals. Now if one uses a local RI metric, such as the overlap[6] or attenuated Coulomb metric,[7,41–43] the number of elements of the three-center tensor $\tilde{\mathbf{B}}$ will grow only linearly with the system size, allowing for an efficient calculation of $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$. Here we want to note that the contraction of $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$ with the RI-matrix and other operations on the $N_{\mathrm{Aux}} \times N_{\mathrm{Aux}}$ matrices $\mathbf{Q}(\mathrm{i}\omega)$ and $\mathbf{Y}(\mathrm{i}\omega)$ (matrix multiplications and eigendecompositions) have a very low prefactor and will therefore only become dominant for exceedingly large molecules,[6] which means that the calculation of $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$ determines the effective scaling behavior.

The AO formulation of $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$ allows for an efficient integral-direct calculation. Therefore, we first carry out the double-Laplace transform to rewrite $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$ as

$$\tilde{Y}_{PQ}(\mathrm{i}\omega) = \sum_{\mu',\nu',\lambda',\sigma'} \tilde{M}^P_{\mu'\nu'}(\mathrm{i}\omega)K_{\mu'\nu',\lambda'\sigma'}\tilde{M}^Q_{\lambda'\sigma'}(\mathrm{i}\omega), \quad (20)$$

with

$$\tilde{M}^P_{\mu'\nu'}(\mathrm{i}\omega) = \sum_{\mu,\nu}\sum_\tau -2w_\tau\cos(\omega\tau)\underline{P}^\tau_{\mu\mu'}\tilde{B}^P_{\mu\nu}\overline{P}^\tau_{\nu\nu'}. \quad (21)$$

Since the occupied and unoccupied pseudodensities are sparse quantities, the number of elements of the three-index quantity $\tilde{\mathbf{M}}(\mathrm{i}\omega)$ also grows only linearly with system size and is efficiently calculated using sparse algebra. Now $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$ can be calculated in a linear scaling fashion by an integral-direct contraction of the AO ERI's with one of the $\tilde{\mathbf{M}}(\mathrm{i}\omega)$ tensors followed by a matrix product over the remaining AO indices,

$$K^Q_{\mu'\nu'}(\mathrm{i}\omega) = \sum_{\lambda',\sigma'} K_{\mu'\nu',\lambda'\sigma'}\tilde{M}^Q_{\lambda'\sigma'}(\mathrm{i}\omega), \quad (22)$$

$$\tilde{Y}_{PQ}(\mathrm{i}\omega) = \sum_{\mu',\nu'} \tilde{M}^P_{\mu'\nu'}(\mathrm{i}\omega)K^Q_{\mu'\nu'}(\mathrm{i}\omega). \quad (23)$$

Linear scaling is achieved by the realization that the locality of the $\tilde{\mathbf{M}}(\mathrm{i}\omega)$ tensor and the locality of AO basis function pairs within the ERI tensor renders only a linear number of ERIs significant. These are then contracted with $\tilde{\mathbf{M}}(\mathrm{i}\omega)$ by a LinK type scheme.[44,45] To identify the significant ERIs, we use Schwarz estimates $Q_{\mu\nu} = (\mu\nu|\mu\nu)^{1/2}$ weighted with the corresponding $\tilde{M}^{max}_{\mu'\nu'}(\mathrm{i}\omega)$ element, where *max* denotes the maximum absolute value over all auxiliary indices, and discard insignificant integrals according to a fixed threshold ϑ,

$$|\tilde{M}^P_{\mu'\nu'}(\mathrm{i}\omega)(\mu'\sigma'|\lambda'\nu')| \leq |\tilde{M}^{max}_{\mu'\nu'}(\mathrm{i}\omega)|Q_{\mu'\sigma'}Q_{\lambda'\nu'} < \vartheta. \quad (24)$$

The remaining significant integrals are contracted with a constant number of $\tilde{M}^P_{\mu'\nu'}(\mathrm{i}\omega)$ elements, determined by the block sparse algebra implementation. In practice, we perform the contraction in Eq. (22) for batches of auxiliary indices to store all necessary quantities in computer memory and to allow for fine grained screening. Per default, we set up the batching scheme by performing the contraction for all auxiliary indices belonging to one atom, where *max* then refers to the batch maximum. We note that the idea of an integral-direct computation of $\mathbf{K}$ along with an RI-decomposed Hartree kernel was first mentioned in the supplementary information of Ref. 18. Through the use of a local RI-metric in conjunction with a transformed non-interacting polarization propagator and the use of the LinK scheme, we arrive at an efficient scheme to calculate beyond RPA correlation energies, which we will refer to as AO-LinK in Sec. IV.

C. Cholesky decomposed density (CDD) RI formulation

The method presented above will scale efficiently for large molecules due to the sparsity of AO quantitites. An issue with AO-based methods is, however, their steep computational scaling with respect to the basis set size N_{Bas} due to the redundancy present in typical AO basis sets, as compared to canonical methods. A solution to this problem has been to exploit the rank deficiency of AO quantities by using pivoted Cholesky decomposition (CD).[7,37,46–50] Pivoted Cholesky decomposition allows decomposing a positive semi-definite matrix into a product of a lower and upper triangular matrix,

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T. \tag{25}$$

The Cholesky factors $\mathbf{L}$ have the same number of rows as the original matrix $\mathbf{A}$ and $rank(\mathbf{A})$ columns. Since the rank of the occupied one-particle density matrix $\mathbf{P}_{\mathrm{occ}}$ equals the number of occupied orbitals, CD allows for huge computational savings when large basis sets are employed. The rank of the virtual one particle density matrix $\mathbf{P}_{\mathrm{virt}}$ equals the number of virtual orbitals. Since especially for large basis sets N_{virt} is of similar size as N_{Bas}, the computational benefits of decomposing $\mathbf{P}_{\mathrm{virt}}$ are not as high. Furthermore, decomposing $\mathbf{P}_{\mathrm{virt}}$ has been found to interfere negatively with matrix sparsity,[7] which is why we omit the decomposition of $\mathbf{P}_{\mathrm{virt}}$ in this work. For a more detailed description of pivoted CD in the context of one particle density matrices, see, e.g., Ref. 7. While the rank discussion also applies to occupied and virtual pseudodensities (where the rank is sometimes even lower than N_{occ} and N_{virt}, respectively), we found it to be more efficient to introduce Cholesky decomposed densities (CDDs) into the calculation of $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$, by decomposing the one-particle density matrix $\mathbf{P}_{\mathrm{occ}} = \mathbf{L}\mathbf{L}^T$ in conjunction with the following equality for occupied pseudodensities:

$$\underline{\mathbf{P}}^{\tau} = \underline{\mathbf{P}}^{\tau}\mathbf{S}\mathbf{P}_{\mathrm{occ}}, \tag{26}$$

where $\mathbf{S}$ is the AO overlap matrix. Thus, we can rewrite Eq. (20) by also inserting the RI for the remaining ERI as

$$\tilde{Y}_{PQ}(\mathrm{i}\omega) = \sum_{\underline{i},\underline{j},\nu',\sigma'} \sum_{R} \tilde{M}^{P}_{\underline{i}\nu'}(\mathrm{i}\omega)\tilde{B}^{R}_{\underline{i}\sigma'}\overline{B}^{R}_{\underline{j}\nu'}\tilde{M}^{Q}_{\underline{j}\sigma'}(\mathrm{i}\omega). \tag{27}$$

Here $\underline{i},\underline{j}$ denote the Cholesky vectors of $\mathbf{P}_{\mathrm{occ}}$, $\tilde{B}^{R}_{\underline{i}\sigma'} = \sum_{\mu'} L^{T}_{\underline{i}\mu'}\tilde{B}^{R}_{\mu'\sigma'}$, $\overline{\mathbf{B}} = \mathbf{C}\tilde{\mathbf{B}}$, and

$$\tilde{M}^{P}_{\underline{i}\nu'}(\mathrm{i}\omega) = \sum_{\tau} -2w_{\tau}\cos(\omega\tau)(\mathbf{L}^{T}\mathbf{S}\underline{\mathbf{P}}^{\tau}\tilde{\mathbf{B}}^{P}\overline{\mathbf{P}}^{\tau})_{\underline{i}\nu'}. \tag{28}$$

Note that we used an asymmetric RI in Eq. (27). In this way, it is possible to significantly reduce the cost of forming $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$ since the number of significant elements of $\tilde{\mathbf{B}}$ scales linearly with system size as opposed to $\mathbf{B}$ and $\overline{\mathbf{B}}$, when a local RI metric is employed. To optimally exploit the locality of the quantities in Eq. (27), the order of summation is crucial. Optimizing the order of summation, subject to formal $\mathcal{O}(\mathcal{N}^3)$ memory requirements and lowest formal scaling behavior, in conjunction with efficient usage of matrix sparsity, we arrive at the sequence shown in the pseudocode in Algorithm 1. Obviously sparse algebra routines are required for efficiency.

Algorithm 1. Pseudocode to optimally exploit sparsity in the formation of $\tilde{\mathbf{Y}}(\mathrm{i}\omega)$ using the CDD-RI formulation.

```
function Calculate Ỹ(iω)
  for R do
    X^{RQ}_{ij}(iω) = Σ_{σ'} B̃^R_{iσ'} M̃^Q_{jσ'}(iω)
    K^Q_{iν'}(iω)+ = Σ_j B̄^R_{jν'} X^{RQ}_{ij}(iω)
  end for
  Ỹ_{PQ}(iω) = Σ_{i,ν'} M̃^P_{iν'}(iω) K^Q_{iν'}(iω)
end function
```

We will refer to these methods as CDD-RI in Sec. IV.

D. Schwarz screened CDD-RI formulation

Memory layouts and caching of modern computers advocate for rather large block dimensions in the blocked sparse algebra routines which need to be traded off against the granularity at which the sparsity in the AO-CDD quantities can be captured. This means that with the most efficient large block sizes, an algorithm exploiting only block sparsity will translate into significant computational savings only for rather large molecules. Since the formal scaling behavior of the method in Sec. II C is $\mathcal{O}(\mathcal{N}^2_{\mathrm{Aux}}\mathcal{N}^2_{\mathrm{occ}}\mathcal{N}_{\mathrm{Bas}})$, this would result in a noticeable prefactor. Screening of individual integrals allows circumventing this problem, and the derivation of the necessary integral estimates is given in the following. The method in this section is currently only applicable to the RPA-SOSEX and RPA-AXK correction and not to RPA-eh-TDHF. We show the derivation here explicitly for RPA-SOSEX.

Using the rotatory invariance of the trace and defining

$$\mathbf{W}(\mathrm{i}\omega) = \mathbf{Q}^{-1}(\mathrm{i}\omega)\log(\mathbf{1} - \mathbf{Q}(\mathrm{i}\omega))\mathbf{Q}^{-1}(\mathrm{i}\omega) + \mathbf{Q}^{-1}(\mathrm{i}\omega). \tag{29}$$

Equation (10) can be rewritten as

$$E_{\mathrm{c}}^{\text{RI-SOSEX}} = -\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\,\mathrm{Tr}\big\{\mathbf{W}(\mathrm{i}\omega)\mathbf{Y}\big\}. \tag{30}$$

Insertion of Eqs. (18) and (27) then yields

$$E_{\mathrm{c}}^{\text{RI-SOSEX}} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\Big\{\sum_{P,Q} M^{P}_{\underline{i},\nu'}(\mathrm{i}\omega)M^{P}_{\underline{j},\sigma'}(\mathrm{i}\omega)B^{Q}_{\underline{i},\sigma'}B^{Q}_{\underline{j},\nu'}\Big\}, \tag{31}$$

with the definition

$$\mathbf{M}(\mathrm{i}\omega) = (-\mathbf{W}(\mathrm{i}\omega))^{1/2}\mathbf{C}^{1/2}\tilde{\mathbf{M}}(\mathrm{i}\omega)\,. \tag{32}$$

Note that we have absorbed the minus sign into the matrix root since $\mathbf{W}(\mathrm{i}\omega)$ is negative definite and this therefore allows for a real-valued, symmetric decomposition.

While Eq. (31) is our final RI-formulation, it is insightful to stress that whereas

$$\sum_{Q} B^{Q}_{\underline{i},\sigma'}B^{Q}_{\underline{j},\nu'} = K_{\underline{i}\nu',\underline{j}\sigma'} \tag{33}$$

is the RI-decomposition of the exchange-type electron repulsion integral, we introduce the RI-decomposition of the coupling-strength averaged screened Coulomb interaction $\overline{W}(\mathrm{i}\omega)$,

$$-\sum_{P} M^{P}_{\underline{i},\nu'}(\mathrm{i}\omega) M^{P}_{\underline{j},\sigma'}(\mathrm{i}\omega) = (\underline{i}\nu'(\mathrm{i}\omega)|\overline{W}(\mathrm{i}\omega)|\underline{j}\sigma'(\mathrm{i}\omega)). \quad (34)$$

This can be easily seen by following the treatment of Secs. II B and II C again, starting from an alternative resummation of Eq. (3) given in Ref. 20,

$$E_{\mathrm{c}}^{\mathrm{SOSEX}} = -\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\Big\{(ia|\overline{W}(\mathrm{i}\omega)|jb)(ib|ja) \times \Pi_0(\mathrm{i}\omega)_{ia,ia}\Pi_0(\mathrm{i}\omega)_{jb,jb}\Big\}, \quad (35)$$

which results in

$$E_{\mathrm{c}}^{\mathrm{SOSEX}} = -\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\mathrm{Tr}\Big\{(\underline{i}\nu'(\mathrm{i}\omega)|\overline{W}(\mathrm{i}\omega)|\underline{j}\sigma'(\mathrm{i}\omega))(\underline{i}\sigma'|\underline{j}\nu')\Big\}. \quad (36)$$

Here we note that not only the coupling-strength averaged screened Coulomb interaction but also the charge distributions $(\underline{i}\nu'(\mathrm{i}\omega)|$ are frequency dependent. An RI-decomposition of the screened Coulomb interaction has been introduced in Ref. 51 and was applied to RPA-SOSEX using a numerical coupling-strength integration in Ref. 20. The approach presented in our work, Eqs. (31) and (29), can in contrast be evaluated directly from the already integrated $\mathbf{Q}(\mathrm{i}\omega)$ and $\mathbf{B}$.

Equation (31) bears close resemblance to the exchange-type term of RI-CDD MP2, where an efficient evaluation technique has been proposed.[37] Due to the exchange type coupling of the local charge distributions, one can select a linear number of Coulomb ERIs [Eq. (33)] and coupling-strength averaged screened Coulomb ERIs [Eq. (34)], which contribute significantly to the final energy. These integrals are then calculated using the RI and directly summed up for the final energy according to Eq. (31).

To find this linear number of significant integrals, one needs to find efficient integral estimates to reduce the number of integrals that explicitly need to be evaluated. While in Ref. 37, QQR-type integral estimates were used, we presently use simple Schwarz estimates since the exchange coupling ensures linear scaling of the number of integrals with respect to the system size,

$$|(\underline{i}\sigma'|\underline{j}\nu')| \leq (\underline{i}\sigma'|\underline{i}\sigma')^{1/2}(\underline{j}\nu'|\underline{j}\nu')^{1/2} = Q_{\underline{i}\sigma'}Q_{\underline{j}\nu'}, \quad (37)$$

$$\begin{aligned}|(\underline{i}\nu'(\mathrm{i}\omega)|\overline{W}(\mathrm{i}\omega)|\underline{j}\sigma'(\mathrm{i}\omega))| &\leq (-(\underline{i}\nu'(\mathrm{i}\omega)|\overline{W}(\mathrm{i}\omega)|\underline{i}\nu'(\mathrm{i}\omega)))^{1/2} \\ &\quad\times(-(\underline{j}\sigma'(\mathrm{i}\omega)|\overline{W}(\mathrm{i}\omega)|\underline{j}\sigma'(\mathrm{i}\omega)))^{1/2} \\ &=: Q_{\underline{i}\nu'}(\mathrm{i}\omega)Q_{\underline{j}\sigma'}(\mathrm{i}\omega), \end{aligned} \quad (38)$$

and hence

$$\begin{aligned}&|(\underline{i}\nu'(\mathrm{i}\omega)|\overline{W}(\mathrm{i}\omega)|\underline{j}\sigma'(\mathrm{i}\omega))(\underline{i}\sigma'|\underline{j}\nu')| \\ &\quad\leq Q_{\underline{i}\nu'}(\mathrm{i}\omega)Q_{\underline{j}\sigma'}(\mathrm{i}\omega)Q_{\underline{i}\sigma'}Q_{\underline{j}\nu'}\,. \end{aligned} \quad (39)$$

This method not only allows for fine grained use of sparsity but also reduces the formal scaling by a factor of $N_{\mathrm{Aux}}/N_{\mathrm{Bas}}$, compared to the method in Sec. II C, to $\mathcal{O}(\mathcal{N}_{\mathrm{Aux}}\mathcal{N}_{\mathrm{occ}}^2\mathcal{N}_{\mathrm{Bas}}^2)$.

While all three-index quantities in Eq. (31) contain the matrix square root of the RI-matrix, $\mathbf{C}^{1/2}$, and therefore linear scaling of the number of significant elements in the $\mathbf{B}$ and $\mathbf{M}(\mathrm{i}\omega)$ tensors is lost, it is still advisable to use a local metric, such as the overlap or attenuated Coulomb metric. This allows using sparsity in the formation of $\mathbf{B}$ and $\mathbf{M}(\mathrm{i}\omega)$, by first forming $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{M}}(\mathrm{i}\omega)$ and then performing the contraction with the respective RI matrix. In the asymptotic limit, this should show quadratic scaling behavior since the local metric ensures a constant number of AO/Cholesky-indices for each auxiliary basis index.

For RPA-AXK, the only difference is that the $\mathbf{W}(\mathrm{i}\omega)$ matrix is given as

$$\begin{aligned}\mathbf{W}(\mathrm{i}\omega) = \mathbf{Q}^{-1}(\mathrm{i}\omega)\log(\mathbf{1}-\mathbf{Q}(\mathrm{i}\omega))\mathbf{Q}^{-1}(\mathrm{i}\omega) \\ +\mathbf{Q}^{-1}(\mathrm{i}\omega)(\mathbf{1}-\mathbf{Q}(\mathrm{i}\omega))^{-1}\end{aligned} \quad (40)$$

and we will refer to these methods as QQ-CDD-RI in Sec. IV.

III. COMPUTATIONAL DETAILS

We have implemented all methods including the canonical formulation and the canonical RI formulation in the FermiONs++[52,53] program package developed in our group. For the canonical RI formulation, we follow the idea given in the supplementary material of Ref. 29 which shows a $\mathcal{O}(\mathcal{N}_{\mathrm{Aux}}^3\mathcal{N}_{\mathrm{occ}}^2)$ scaling behavior and makes optimal use of efficient linear algebra libraries. For the canonical RI formulation, we use the Coulomb RI-metric. Unless noted otherwise, we employ the attenuated Coulomb metric with $\omega = 0.1$ for all low scaling algorithms, as has been shown to be optimal for dRPA calculations.[7]

For the grids of the numerical frequency integration and the double-Laplace transform, we employ the minimax grids presented in Ref. 14. The rationale behind the fitting procedure given in Ref. 14 is easily extendable to the beyond RPA methods treated here. For RPA-SOSEX and RPA-eh-TDHF, the second-order approximation corresponds to the MP2 energy,[29] as compared to the direct (opposite spin) MP2 energy for dRPA. Since this only affects the integral tensors but leaves the integrand unchanged, the exact same procedure as for dRPA can be used. As discussed in the supplementary information of Ref. 18, RPA-SOSEX and RPA-AXK recover the same second-order contribution, which is why the grids are also suited for RPA-AXK.

We employ 15 grid points for the time and frequency grid in the general case. Where the integration interval is sufficiently small for 15 grid points not to yield any more accuracy in our double precision implementation, less grid points are automatically employed.

As a sparse algebra format, we use the blocked compressed sparse row (BCSR) format optimized for quantum chemical methods.[54,55] We employ a block size of 50 elements and a sparsity threshold of 10^{-7} unless noted otherwise. For the canonical algorithms, we of course always employ dense algebra. For the remaining algorithms, we will mention this explicitly in Sec. IV. The integral threshold ϑ for the AO-LinK formulation was set to 10^{-10}. The screening threshold for the QQ-CDD-RI method was set to 10^{-9}. The Schwarz estimates used for the QQ-CDD-RI formulation are calculated using the RI as in Ref. 37.

The exact procedure for pivoted CD of density matrices is detailed in, e.g., Refs. 7 and 37. For the generation of the pseudodensities, we employ a Fermi shift to enhance numerical stability as detailed in Ref. 56.

The evaluation of the RI-RPA-eh-TDHF correlation energy using Eq. (12) with given $\mathbf{Q}(i\omega)$ and $\mathbf{Y}(i\omega)$ poses some numerical challenges due to the need to explicitly calculate $\mathbf{Q}^{-1}(i\omega)$, as $\mathbf{Q}(i\omega)$ can turn out to be numerically close to a singular matrix. Therefore, we follow the idea presented in Ref. 40 and project all quantities on the space spanned by the eigenvectors of $\mathbf{Q}(i\omega)$, corresponding to non-zero eigenvalues (in this case absolute values larger than 10^{-14}). Furthermore, the argument in the logarithm is non-symmetric and therefore complex eigenvectors are possible. A numerically more stable version of Eq. (12) is work in progress.

All calculations use reference KS orbitals obtained from DFT calculations with the Perdew-Burke-Ernzerhof (PBE) functional.[57] We employ def2-SVP, def2-TZVP, and def2-QZVP basis sets[58,59] along with their corresponding RI basis sets.[60,61] Total energies are obtained by adding correlation energies to the Hartree-Fock energy evaluated with the KS-orbitals. All calculations employ the frozen core approximation.

IV. RESULTS AND DISCUSSION

Since the purpose of this work is to show the efficient yet accurate calculation of $\mathbf{Y}(i\omega)$, we present our results exemplarily using RPA-SOSEX correlation energies. We will start with showing the accuracy of our methods using the S22 test set,[62] before demonstrating the efficiency of our methods. We finish with an illustrative large scale application, by presenting RPA-SOSEX results for the L7 test set of large, dispersion dominated molecules.[63]

A. Convergence of the frequency quadrature and the double-Laplace transform

The convergence of the frequency quadrature and the double-Laplace transform has been shown nicely for dRPA correlation energies in a plane wave implementation.[14] Here we want to show briefly that minimax grids also allow for well-converged results for beyond RPA correlation energies with less than 20 grid points. Therefore, we calculated the RPA-SOSEX correlation energies for the methane monomer and dimer of the S22 test set[62] using the def2-TZVP and def2-QZVP basis sets, along with the corresponding RI basis set. As a reference, we use canonical RI-RPA-SOSEX results obtained with Clenshaw-Curtis quadrature with the RPA adjusted parameter optimization,[9] which has been shown to also work for closed-shell RPA-SOSEX.[29] For the reference calculations, we employ 500 grid points, which ensures well-converged results.

Table I shows that the minimax grids allow converging the absolute correlation energies to below 10^{-10} hartree with respect to the numerical frequency integration. When additionally using the corresponding grids for the double-Laplace transform of the non-interacting polarization propagator, the minimax grids allow converging the total energies to 10^{-10} hartree as well (Table II). Beyond 13 grid points, the interaction energy fluctuates in the range of 10^{-6} kcal/mol, which is well beyond the accuracy of the methods under inspection. Based on these results, we have chosen 15 grid points in Secs. IV B–IV D.

TABLE I. Convergence of the numerical frequency integration using minimax grids referenced against a well converged Clenshaw-Curtis quadrature with 500 grid points. All results are RI-RPA-SOSEX correlation energies of the methane monomer/dimer from the S22 test set.[62] The interaction energy is denoted as ΔE.

	def2-TZVP		
No. pts.	Monomer (hartree)	Dimer (hartree)	ΔE (kcal/mol)
10	−0.201 615 930 3	−0.404 493 209 8	−0.304 712
13	−0.201 615 931 6	−0.404 493 214 3	−0.304 713
15	−0.201 615 931 6	−0.404 493 214 4	−0.304 713
18	−0.201 615 931 6	−0.404 493 214 4	−0.304 713
500	−0.201 615 931 6	−0.404 493 214 4	−0.304 713
	def2-QZVP		
No. pts.	Monomer (hartree)	Dimer (hartree)	ΔE (kcal/mol)
10	−0.226 027 242 2	−0.453 494 471 6	−0.377 841
13	−0.226 027 338 4	−0.453 494 608 6	−0.377 806
15	−0.226 027 339 7	−0.453 494 612 4	−0.377 807
18	−0.226 027 339 9	−0.453 494 613 0	−0.377 807
500	−0.226 027 339 9	−0.453 494 613 0	−0.377 807

B. Calculations on the S22 test set

To further examine the accuracy of all our presented methods, we have performed calculations on the entire S22 test set.[62] We have performed calculations using def2-SVP, def2-TZVP, and def2-QZVP basis sets along with their

TABLE II. Convergence of the numerical frequency integration and double-Laplace transform of the non-interacting polarization propagator using minimax grids referenced against a well converged Clenshaw-Curtis quadrature with 500 grid points. The reference results are RI-RPA-SOSEX correlation energies, while the convergence is tested using CDD-RI-RPA-SOSEX (Coulomb metric) correlation energies. All correlation energies are calculated for the methane monomer/dimer from the S22 test set.[62] The interaction energy is denoted as ΔE.

	def2-TZVP		
No. pts.	Monomer (hartree)	Dimer (hartree)	ΔE (kcal/mol)
10	−0.201 615 914 1	−0.404 493 171 7	−0.304 708
13	−0.201 615 931 3	−0.404 493 213 8	−0.304 713
15	−0.201 615 931 6	−0.404 493 214 4	−0.304 713
18	−0.201 615 931 6	−0.404 493 214 4	−0.304 713
500	−0.201 615 931 6	−0.404 493 214 4	−0.304 713
	def2-QZVP		
No. pts.	Monomer (hartree)	Dimer (hartree)	ΔE (kcal/mol)
10	−0.226 028 705 4	−0.453 497 574 2	−0.377 951
13	−0.226 027 299 6	−0.453 494 532 3	−0.377 806
15	−0.226 027 332 9	−0.453 494 596 5	−0.377 799
18	−0.226 027 339 6	−0.453 494 612 4	−0.377 807
500	−0.226 027 339 9	−0.453 494 613 0	−0.377 807

corresponding RI-basis sets. As a reference, we have performed calculations using both the canonical implementation and the RI-canonical implementation using the Coulomb metric.

The error introduced through the use of the RI within the S22 test set as compared to the canonical implementation is 0.014 and 0.009 kcal/mol in the mean absolute error (MAE) for the def2-SVP and def2-TZVP basis set, respectively, and is therefore negligible (Fig. 1). For the def2-QZVP results, we therefore employ the canonical RI implementation as a reference for our newly developed methods since the computational cost and memory requirements for the canonical implementation for the def2-QZVP basis set become unfeasible.

To examine the impact of the different approximations made here, we perform calculations using separately the numerical double-Laplace transform with the Coulomb RI-metric and calculations using the numerical double-Laplace transform with the attenuated Coulomb metric also employing sparse algebra. The results for the def2-SVP and def2-TZVP basis sets are shown in Fig. 1. The results for the def2-QZVP basis are shown in Fig. 2.

For all basis sets employed here, the error introduced through the double-Laplace transform is barely noticeable, as can be seen from the error bars for the CDD-RI and QQ-CDD-RI calculations, as compared to the plain RI error. The error introduced by changing from the Coulomb to the attenuated Coulomb metric shows a maximum of 0.004 kcal/mol for the def2-SVP basis in the MAE as compared to the RI-canonical value and is therefore also negligible. For the QQ-CDD-RI formulation, the maximum deviation in the MAE as compared to the RI-canonical implementation is 0.004 kcal/mol, showing that the Schwarz screening does not introduce a relevant error. The maximum error caused by Schwarz screening and the use of the attenuated Coulomb metric combined is 0.005 kcal/mol for the def2-QZVP basis set. The use of sparse algebra with the thresholds given in Sec. III in conjunction with the def2-QZVP basis for the CDD-RI formulation leads to significantly higher errors for some individual systems and a total MAE of 0.3 kcal/mol. Employing dense algebra with the attenuated Coulomb metric as shown in Fig. 2 produces again a negligible deviation, as compared to the RI-canonical implementation. This problem is related to the atomic blocking used in BCSR, which would require larger block sizes and a tighter sparsity threshold for large basis sets.

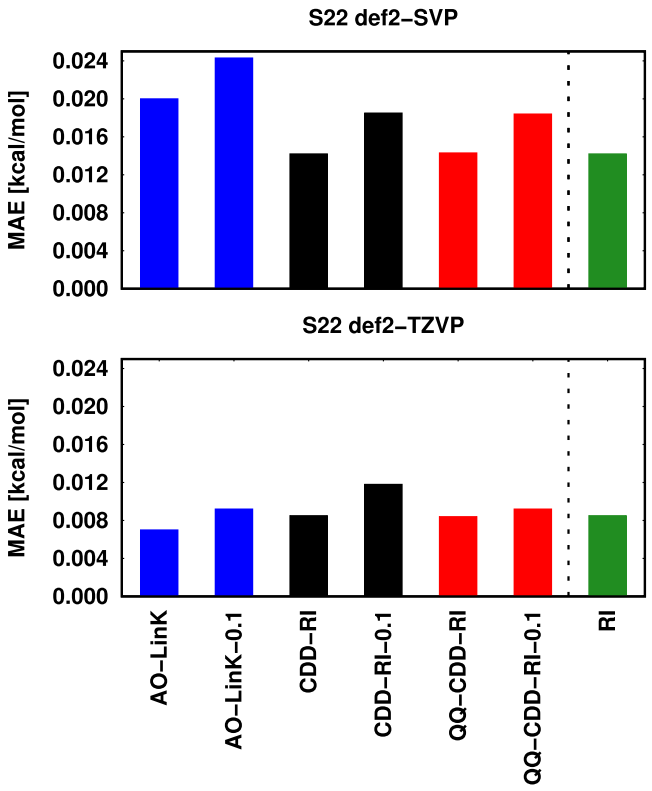


FIG. 1. Bar chart showing the mean absolute error of the RPA-SOSEX interaction energies of the S22 test set as compared to the canonical implementation using def2-SVP and def2-TZVP basis sets along with the corresponding RI-basis set. As RI-metric the Coulomb metric is employed unless the suffix 0.1 is appended, which implies use of the attenuated Coulomb metric with $\omega = 0.1$.

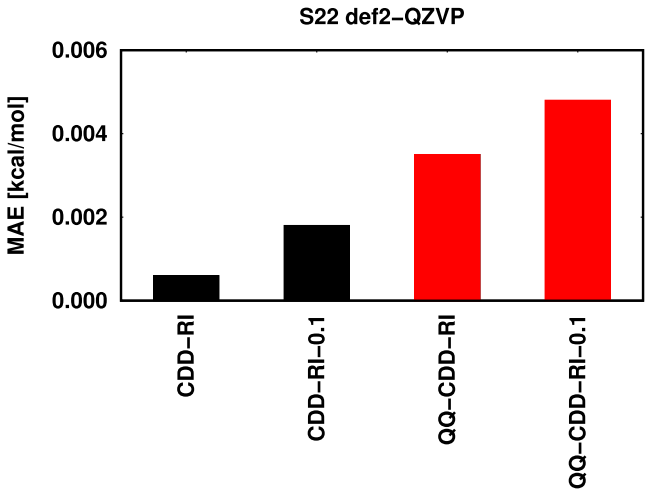


FIG. 2. Bar chart showing the mean absolute error of the RPA-SOSEX interaction energies of the S22 test set as compared to the RI-canonical implementation using the def2-QZVP basis set along with the corresponding RI-basis set. As RI-metric the Coulomb metric is employed unless the suffix 0.1 is appended, which implies use of the attenuated Coulomb metric with $\omega = 0.1$. For the QQ-CDD-RI, CDD-RI and CDD-RI-0.1 results we have employed dense algebra, while the QQ-CDD-RI-0.1 results were produced using sparse-algebra.

For the AO-LinK method with the chosen thresholds, the deviation in the MAE shows a maximum of 0.020 kcal/mol (0.024 kcal/mol with the attenuated Coulomb metric) for the def2-SVP basis set. This additional error as compared to the CDD-RI variants could stem from the integral screening employed in the integral-direct computation. We omitted the computation of S22 interaction energies with the method formulated in a pure AO basis for the def2-QZVP basis set due to too high computational cost. The problem here stems from the aforementioned bad scaling of pure AO methods with respect to the basis set size due to the redundancy present in pure AO basis sets. This already shows one advantage of the two low scaling methods, which employ CDD to avoid this problem and show similar formal scaling as the RI-canonical implementation, while still being efficient for large molecular systems, as will be shown in Secs. IV C and IV D.

To put the above errors into perspective, the MAE of the canonical implementation using def2-SVP and def2-TZVP

basis sets is 0.85 kcal/mol and 0.35 kcal/mol, respectively, and 0.27 kcal/mol for the def2-QZVP basis using the RI-canonical implementation (all values referenced against the S22A revised results of Ref. 64). This shows that the errors introduced through our approximations are at least one order of magnitude smaller than the method error calculated with the reference implementation and therefore insignificant.

The mean absolute percentage error of 7.0% for the largest def2-QZVP basis set as compared to the S22A revised reference results[64] is in good agreement with the values of 9.5% and 10.5% reported for a plane wave[34] and numerical atomic orbital implementation[20] of RPA-SOSEX, respectively.

C. Efficiency and asymptotic scaling behavior

To show the efficiency and the low asymptotic scaling behavior of our presented methods for large systems with a nonvanishing HOMO-LUMO gap, we performed RPA-SOSEX correlation energy calculations for linear alkanes up to $C_{160}H_{322}$. As has been discussed, e.g., in Ref. 6, these systems serve as good and practical systems to determine the asymptotic scaling of a quantum chemical method. We have performed all calculations using the def2-SVP basis with the corresponding RI basis set. We compare the scaling of our method to the RI-canonical implementation. The calculations were performed using 12 threads on dual-processor Intel Xeon CPU E5-2620 machines with 64 GB of memory. All timings shown in this section correspond to the wall time needed to perform the beyond RPA correction. The time required for the calculation of $\mathbf{Q}(i\omega)$ is excluded since we want to focus on the efficient calculation of the beyond RPA energy and $\mathbf{Q}(i\omega)$ is part of the dRPA calculation. For the performance of the linear-scaling dRPA implementation, see Refs. 7 and 8. As a rough estimate about the comparative cost to form $\mathbf{Q}(i\omega)$ vs. the cost to form $\mathbf{Y}(i\omega)$, we consider the calculation of $C_{160}H_{322}$ for which the calculation of $\mathbf{Q}(i\omega)$ is presently faster by a factor of about 70. This can be explained with the fact that the formal scaling of the linear scaling CDD-RI-dRPA algorithm with respect to the molecular size is smaller by one power, leading to a significantly smaller prefactor.

To calibrate the sparsity settings, i.e., thresholds and block sizes for our sparse algebra routines, we compare the results for $C_{40}H_{82}$ against the RI-canonical implementation. Using the thresholds and block sizes listed in Sec. III, this leads to an error of 57 and 42 μhartree for $C_{40}H_{82}$ in the absolute RPA-SOSEX correlation energy for the CDD-RI and QQ-CDD-RI formulation, respectively. For the AO-LinK method with the chosen thresholds, the deviation is 2.9 *m*hartree. While this deviation is significantly higher, one has to keep in mind that in this method the integral tensor corresponding to $\mathbf{K}$ is computed without the RI approximation, which means that the canonical RI calculation contains the additional RI error introduced through the RI decomposition of $\mathbf{K}$.

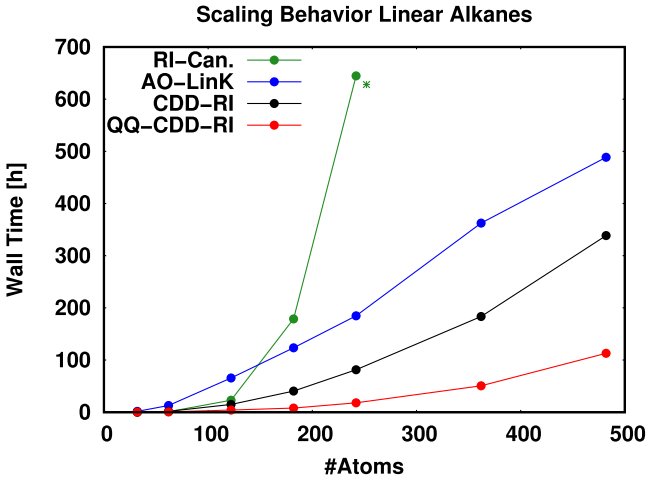


FIG. 3. Timings for the calculations on a set of linear alkanes using the def2-SVP basis set with the corresponding RI basis set. The AO-LinK, CDD-RI, and QQ-CDD-RI methods employ the attenuated Coulomb metric with $\omega = 0.1$ along with sparse algebra. The point labelled with an asterisk was estimated conservatively based on the timing for a subset of the frequency points.

The wall times shown in Fig. 3 for RPA-SOSEX calculations on a series of alkanes from $C_{10}H_{22}$ to $C_{160}H_{322}$ exemplify that all our presented methods significantly outperform the RI-canonical implementation for large, electronically sparse systems in terms of computational efficiency: The AO-LinK method shows a crossover to the RI-canonical method at $C_{60}H_{122}$, the CDD-RI method at $C_{40}H_{82}$, and the QQ-CDD-RI method already at $C_{20}H_{42}$. Extrapolating the wall-time for the RI-canonical method for the largest system ($C_{160}H_{322}$) assuming an $\mathcal{O}(\mathcal{N}^5)$ scaling behavior, the speed-up obtained with the AO-LinK method is 42×, with the CDD-RI method is 61×, and with the QQ-CDD-RI method is 183×.

TABLE III. Wall times and observed computational complexities for a set of linear alkanes using the def2-SVP basis set along with the corresponding RI basis set for the different methods presented and the RI canonical implementation. The computational complexities were calculated using the preceding calculation in the table. The number labelled with an asterisk was estimated conservatively based on the timing for a subset of the frequency points.

	AO-LinK		CDD-RI		QQ-CDD-RI		RI-canonical	
Atoms	Time (h)	Scaling	Time (h)	Scaling	Time (h)	Scaling	Time (h)	Scaling
62	12.8		1.7		0.6		0.9	
122	65.6	2.4	14.9	3.2	4.1	2.8	22.9	4.8
242	184.7	1.5	81.4	2.5	18.0	2.1	644.7*	4.9
362	362.4	1.7	183.4	2.0	50.6	2.6		
482	488.6	1.0	338.6	2.1	112.8	2.8		

The observed computational scaling behavior of our methods is examined in more detail in Table III. The RI-canonical implementation shows an $\mathcal{O}(\mathcal{N}^5)$ scaling behavior as expected. For the AO-LinK method, the observed scaling behavior in the limit of sparse molecules is linear; for the CDD-RI method, quadratic; and for the QQ-CDD-RI method, sub-cubic. To show that for the QQ-CDD-RI formulation we indeed only calculate a linear number of significant integrals, we have counted the number of calculated integrals during each of the calculations on the linear alkanes. The result is shown in Fig. 4. As can be seen clearly, the number of integrals shows early on perfect linear scaling.

D. Large scale calculations

In this section, we show that the methods based on CDDs also outperform the canonical-RI variant for larger than double-ζ basis sets. Furthermore, we apply the QQ-CDD-RI method to present RPA-SOSEX results for the L7 test set of large, dispersion dominated molecules[63] with up to triple-ζ basis sets.

First, to study the performance of our methods for larger basis sets, we have performed RPA-SOSEX correlation energy calculations on linear alkanes up to $C_{80}H_{162}$ with the canonical RI implementation and the CDD-RI and QQ-CDD-RI variant using the def2-TZVP basis set. All calculations were performed using 16 threads on a dual-processor Intel Xeon CPU E5-2667 machine. The results presented in Fig. 5 show that both the CDD-RI and the QQ-CDD-RI variant outperform the canonical implementation for large systems. The crossover to the canonical implementation occurs at $C_{60}H_{122}$ with the CDD-RI method and at $C_{40}H_{82}$ with the QQ-CDD-RI method. Furthermore, to show that the QQ-CDD-RI method is beneficial also for quadruple-ζ basis sets, we have performed a RPA-SOSEX correlation energy calculation on the $C_{40}H_{82}$ molecule with the def2-QZVP basis set and the corresponding RI-basis set. Here the calculation of the RPA-SOSEX correlation energy takes 104 h for the QQ-CDD-RI method, while the calculation using the RI-canonical implementation requires 408 h (extrapolated from 8 of the 15 frequency points).

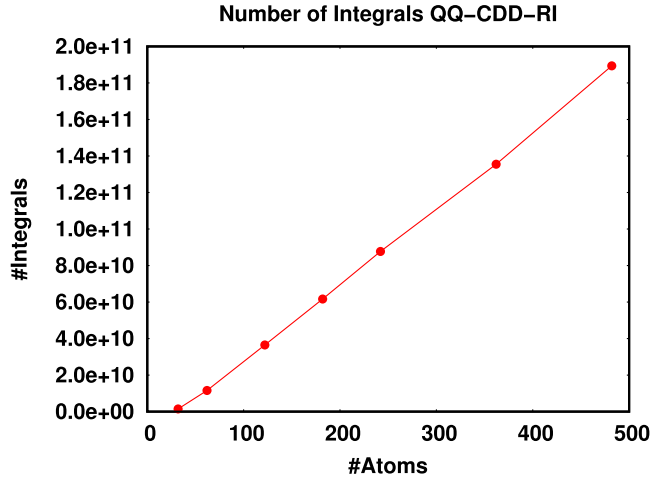


FIG. 4. Number of integrals calculated with the QQ-CDD-RI method for a set of linear alkanes using the def2-SVP basis set with the corresponding RI basis set. The attenuated Coulomb metric was employed with $\omega = 0.1$ along with sparse algebra.

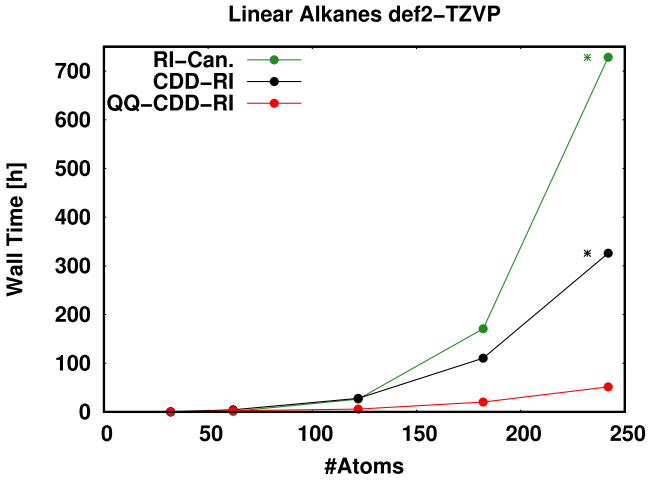


FIG. 5. Timings for the calculations on a set of linear alkanes using the def2-TZVP basis set with the corresponding RI basis set. The CDD-RI, and QQ-CDD-RI methods employ the attenuated Coulomb metric with $\omega = 0.1$ along with sparse algebra. Points labelled with an asterisk were estimated conservatively based on the timings for a subset of the frequency points.

This speed-up of a factor of 3.9 still compares well against the speed-ups obtained with triple-ζ (speed-up: 4.5) and double-ζ (speed-up: 5.6) basis sets obtained with the same method. These results show that also for larger than double-ζ basis sets significant speed-ups over the canonical implementation can be obtained with the CDD-RI and QQ-CDD-RI methods, supported by the dimensionality reduction of the AO basis set via Cholesky decomposition.

Finally, we present RPA-SOSEX benchmark results for the L7 test set[63] with def2-SVP and def2-TZVP basis sets along with their corresponding RI-basis sets. The results shown in Table IV exemplify two important aspects: The first aspect is that our QQ-CDD-RI variant can also treat large molecular systems without excessive sparsity in the electronic structure such as the circumcoronene···guanine-cytosine base pair complex. The second aspect is that including exchange effects yields a significant accuracy gain over plain dRPA for the L7 test set, which can be seen when comparing the results to dRPA results presented in Ref. 7 obtained with the same basis set. The RPA-SOSEX results with double-ζ basis set already improve upon dRPA results with a triple-ζ basis set. The triple-ζ RPA-SOSEX results show a MAE to the reference results[63] of less than 1 kcal/mol, which is the desired chemical accuracy.

TABLE IV. Benchmark results showing the root mean squared deviation (RMSD), mean absolute and mean signed error (MAE and MSE) for the L7 test set[63] of large, dispersion dominated systems as compared to the reference results. The CDD-dRPA values were taken from Ref. 7. The RPA-SOSEX values were calculated using the QQ-CDD-RI approach described in this work with the attenuated Coulomb metric ($\omega = 0.1$) along with sparse algebra.

	RMSD	MAE	MSE
dRPA (def2-TZVP)[7]	2.90	2.45	−2.08
RPA-SOSEX (def2-SVP)	2.44	1.94	−0.13
RPA-SOSEX (def2-TZVP)	1.19	0.81	0.11

V. CONCLUSION

We have introduced three methods that enable efficient beyond RPA calculations for large molecular systems up to 500 atoms, while the accuracy is under full numerical control. These developments significantly extend the applicability of beyond RPA methods by reducing the computational cost compared to the canonical formulation with and without RI. Next to the beyond RPA variants mentioned in the main text, our methods are obviously also applicable to the short-range RPA-SOSEX variant recently introduced by us.[65] For the present range of system sizes, we recommend the QQ-CDD-RI method for general use; since even though it shows the worst asymptotic scaling behavior of the newly presented methods, it is most efficient for a wide range of molecular sizes and shows very good scaling behavior with respect to basis set size. Furthermore, we have shown that the methods employing CDDs also allow for significant computational savings when larger than double-ζ basis sets are used, which ensures the applicability of these methods, since RPA correlation energies require rather large basis sets to obtain converged results.[66] To this end, range separated variants of RPA correlation energies[67–72] would be a valuable addition to our approach. These schemes reduce the basis set dependence significantly so that double-ζ results were shown to be sufficiently accurate already.[73]

ACKNOWLEDGMENTS

The authors thank Dr. A. Luenser (LMU Munich) for helpful discussions. Financial support was provided by the Excellence Cluster EXC114 (CIPSM) and the SFB749 by the Deutsche Forschungsgemeinschaft (DFG). C.O. acknowledges, in addition, financial support as a Max-Planck Fellow at the MPI-FKF Stuttgart.

APPENDIX: DERIVATION FOR RI-RPA-AXK

To derive the RI-expression for RPA-AXK, we expand the logarithm and the inverse of $(\mathbf{1} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V})$ appearing in Eq. (4) as a series

$$\mathrm{Tr}\Big\{\log\left(\mathbf{1} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)\mathbf{V}^{-1}\mathbf{K}\Big\} = \mathrm{Tr}\Big\{-\sum_{n=1}^{\infty}\frac{(\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V})^n\mathbf{V}^{-1}\mathbf{K}}{n}\Big\}, \tag{A1}$$

$$\mathrm{Tr}\Big\{\left(\mathbf{1} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)^{-1}\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\Big\} = \mathrm{Tr}\Big\{\sum_{n=1}^{\infty}(\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V})^{n-1}\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\Big\}. \tag{A2}$$

For $n = 1$, the terms of the two series cancel. Inserting the RI for $\mathbf{V}$ and $\mathbf{K}$, one can introduce the definitions of $\mathbf{Q}(\mathrm{i}\omega)$ and $\mathbf{Y}(\mathrm{i}\omega)$ [see Eqs. (9) and (13)] using the cyclic invariance of the trace,

$$\mathrm{Tr}\Big\{\log\left(\mathbf{1} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)\mathbf{V}^{-1}\mathbf{K} + \left(\mathbf{1} - \mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{V}\right)^{-1}\mathbf{\Pi}_0(\mathrm{i}\omega)\mathbf{K}\Big\} = \mathrm{Tr}\Big\{-\sum_{n=2}^{\infty}\frac{\mathbf{Q}(\mathrm{i}\omega)^{(n-2)}\mathbf{Y}(\mathrm{i}\omega)}{n} + \sum_{n=2}^{\infty}\mathbf{Q}(\mathrm{i}\omega)^{(n-2)}\mathbf{Y}(\mathrm{i}\omega)\Big\}. \tag{A3}$$

This can be brought to the closed form of Eq. (11), where again the first term of each sum cancels.

[1] H. Eshuis, J. E. Bates, and F. Furche, Theor. Chem. Acc. **131**, 1084 (2012).
[2] X. Ren, P. Rinke, C. Joas, and M. Scheffler, J. Mater. Sci. **47**, 7447 (2012).
[3] G. P. Chen, V. K. Voora, M. M. Agee, S. G. Balasubramani, and F. Furche, Annu. Rev. Phys. Chem. **68**, 421 (2017).
[4] F. Furche, Phys. Rev. B **64**, 195120 (2001).
[5] M. Kállay, J. Chem. Phys. **142**, 204105 (2015).
[6] H. F. Schurkus and C. Ochsenfeld, J. Chem. Phys. **144**, 031101 (2016).
[7] A. Luenser, H. F. Schurkus, and C. Ochsenfeld, J. Chem. Theory Comput. **13**, 1647 (2017).
[8] D. Graf, M. Beuerle, H. F. Schurkus, A. Luenser, G. Savasci, and C. Ochsenfeld, J. Chem. Theory Comput. **14**, 2505–2515 (2018).
[9] H. Eshuis, J. Yarkony, and F. Furche, J. Chem. Phys. **132**, 234114 (2010).
[10] N. Shenvi, H. van Aggelen, Y. Yang, and W. Yang, J. Chem. Phys. **141**, 024119 (2014).
[11] E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, J. Chem. Phys. **137**, 044103 (2012).
[12] E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, and T. J. Martínez, J. Chem. Phys. **137**, 221101 (2012).
[13] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, J. Chem. Phys. **137**, 224106 (2012).
[14] M. Kaltak, J. Klimeš, and G. Kresse, J. Chem. Theory Comput. **10**, 2498 (2014).
[15] J. Wilhelm, P. Seewald, M. Del Ben, and J. Hutter, J. Chem. Theory Comput. **12**, 5851 (2016).
[16] J. F. Dobson, J. Wang, B. P. Dinte, K. McLennan, and H. M. Le, Int. J. Quantum Chem. **101**, 579 (2005).
[17] J. F. Dobson and T. Gould, J. Phys. Condens. Matter **24**, 073201 (2012).
[18] J. E. Bates and F. Furche, J. Chem. Phys. **139**, 171103 (2013).
[19] J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse, and M. Scheffler, New J. Phys. **14**, 043002 (2012).
[20] X. Ren, P. Rinke, G. E. Scuseria, and M. Scheffler, Phys. Rev. B **88**, 035120 (2013).
[21] T. M. Henderson and G. E. Scuseria, Mol. Phys. **108**, 2511 (2010).
[22] G. E. Scuseria, T. M. Henderson, and D. C. Sorensen, J. Chem. Phys. **129**, 231101 (2008).
[23] D. Langreth and J. Perdew, Solid State Commun. **17**, 1425 (1975).
[24] D. C. Langreth and J. P. Perdew, Phys. Rev. B **15**, 2884 (1977).
[25] F. Furche, J. Chem. Phys. **129**, 114105 (2008).
[26] A. Grüneis, M. Marsman, J. Harl, L. Schimka, and G. Kresse, J. Chem. Phys. **131**, 154115 (2009).
[27] D. L. Freeman, Phys. Rev. B **15**, 5512 (1977).
[28] J. G. Ángyán, R.-F. Liu, J. Toulouse, and G. Jansen, J. Chem. Theory Comput. **7**, 3116 (2011).
[29] B. Mussard, D. Rocca, G. Jansen, and J. G. Ángyán, J. Chem. Theory Comput. **12**, 2191 (2016).
[30] A. Dixit, J. G. Ángyán, and D. Rocca, J. Chem. Phys. **145**, 104105 (2016).
[31] A. Heßelmann and A. Görling, Mol. Phys. **108**, 359 (2010).
[32] J. Erhard, P. Bleiziffer, and A. Görling, Phys. Rev. Lett. **117**, 143002 (2016).
[33] P. Bleiziffer, A. Heßelmann, and A. Görling, J. Chem. Phys. **136**, 134102 (2012).
[34] A. Dixit, J. Claudot, S. Lebègue, and D. Rocca, J. Chem. Theory Comput. **13**, 5432 (2017).
[35] J. E. Moussa, J. Chem. Phys. **140**, 014107 (2014).
[36] A. Heßelmann, J. Chem. Phys. **146**, 174110 (2017).
[37] S. A. Maurer, L. Clin, and C. Ochsenfeld, J. Chem. Phys. **140**, 224112 (2014).
[38] M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).
[39] H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
[40] H. F. Schurkus, A. Luenser, and C. Ochsenfeld, J. Chem. Phys. **146**, 211106 (2017).
[41] Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, Proc. Natl. Acad. Sci. U. S. A. **102**, 6692 (2005).
[42] Y. Jung, Y. Shao, and M. Head-Gordon, J. Comput. Chem. **28**, 1953 (2007).
[43] S. Reine, E. Tellgren, A. Krapp, T. Kjærgaard, T. Helgaker, B. Jansik, S. Høst, and P. Salek, J. Chem. Phys. **129**, 104101 (2008).

[44]C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[45]C. Ochsenfeld, Chem. Phys. Lett. **327**, 216 (2000).
[46]A. Luenser, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **145**, 124103 (2016).
[47]H. Koch, A. S. de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003).
[48]J. Zienau, L. Clin, B. Doser, and C. Ochsenfeld, J. Chem. Phys. **130**, 204112 (2009).
[49]N. J. Higham, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 251 (2009).
[50]H. Harbrecht, M. Peters, and R. Schneider, Appl. Numer. Math. **62**, 428 (2012).
[51]X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, New J. Phys. **14**, 053020 (2012).
[52]J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013).
[53]J. Kussmann and C. Ochsenfeld, J. Chem. Theory Comput. **11**, 918 (2015).
[54]J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **127**, 054103 (2007).
[55]F. G. Gustavson, ACM Trans. Math. Software **4**, 250 (1978).
[56]P. Y. Ayala and G. E. Scuseria, J. Chem. Phys. **110**, 3660 (1999).
[57]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[58]F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).
[59]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[60]F. Weigend, M. Haser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[61]C. Hättig, Phys. Chem. Chem. Phys. **7**, 59 (2005).
[62]P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[63]R. Sedlak, T. Janowski, M. Pitonak, J. Rezac, P. Pulay, and P. Hobza, J. Chem. Theory Comput. **9**, 3364 (2013).
[64]T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **132**, 144104 (2010).
[65]M. Beuerle and C. Ochsenfeld, J. Chem. Phys. **147**, 204107 (2017).
[66]H. Eshuis and F. Furche, J. Chem. Phys. **136**, 084105 (2012).
[67]J. Toulouse, I. C. Gerber, G. Jansen, A. Savin, and J. G. Ángyán, Phys. Rev. Lett. **102**, 096404 (2009).
[68]W. Zhu, J. Toulouse, A. Savin, and J. G. Ángyán, J. Chem. Phys. **132**, 244108 (2010).
[69]J. Toulouse, W. Zhu, J. G. Ángyán, and A. Savin, Phys. Rev. A **82**, 032502 (2010).
[70]B. G. Janesko, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **130**, 081105 (2009).
[71]B. G. Janesko, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **131**, 034110 (2009).
[72]R. M. Irelan, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **135**, 094105 (2011).
[73]B. Mussard, P. Reinhardt, J. G. Ángyán, and J. Toulouse, J. Chem. Phys. **142**, 154123 (2015).

Chapter 4

Conclusions and Outlook

This thesis presents developments to increase the efficiency and the accuracy of DFT and ACFDT methods. For short-range hybrid DFT calculations screening-methods exploiting the short-range nature of the attenuated Coulomb operator in linear-scaling exchange methods were introduced to significantly reduce the computational requirements. Enabled by these developments, short-range hybrid DFT calculations are now computationally similarly expensive as pure DFT, while including a fraction of exchange for increased accuracy. This establishes short-range hybrid DFT as a lower-cost alternative to conventional hybrid DFT for large molecular systems.
In the field of ACFDT methods, this thesis contributes in various different ways. The optimized transformations and quadratures in combination with the multinode parallel algorithm mature the RPA to a numerically very accurate and truly large-scale method. This is shown by a study on the displacement of two COF layers, which previously was beyond the scope of RPA methods.
Furthermore, the realm of first order RPA analytical gradients is extended from the few atom scale to systems with several hundred atoms. This allows for the first time to perform theoretical studies beyond ground state energy calculations for large and complex systems at the RPA level of theory, which is indispensable in making the RPA a general purpose, non-empirical alternative to conventional DFT.
Moreover, a new RPA with exchange scheme is put forward, which combines the benefits of RPA and RPA-SOSEX correlation energies. This results in a more balanced description across systems prone to self-interaction error and static correlation effects, adding a highly accurate method to the field of RPA with exchange schemes.
Finally, techniques are presented allowing for linear- and low-scaling calculation of RPA with exchange correlation energies. These developments allow to show the superiority of RPA-SOSEX over plain RPA and other correlation methods such as MP2 for large, dispersion dominated systems. In the future, this will enable to further test and apply beyond RPA methods to large and complex chemical systems.
Together these developments represent a significant contribution to the field of ACFDT methods, which will enable to use RPA and RPA with exchange methods in chemical research, opening the way to accurate theoretical insights.
Future projects will aim to reduce the large basis set requirements of ACFDT calculations and circumvent the dependence on the semi-local DFT reference. For the former problem we currently develop the combination of linear- and low-scaling RPA and RPA with exchange methods with range-separated DFT approaches. Here the

short-range correlation effects, which include the electron-electron cusp are described with a short-range density functional, which shows faster convergence with the basis set. The long-range correlation is then accurately described with the ACFDT methods with moderately sized basis sets.
To cure the dependence of ACFDT methods on the reference KS determinant, we currently devise a self-consistent AO-RPA method. Here, an RPA Hamiltonian, defined as the derivative of the RPA energy expression with respect to the one-particle density matrix, is employed. Most of the necessary ingredients for this procedure also appear in the presented low-scaling analytical RPA gradients. Self-consistent RPA will further reduce the dependence of ACFDT methods on conventional DFT.

Bibliography

[1] E. Schrödinger, *Phys. Rev.* **1926**, *28*, 1049–1070.

[2] M. Born, R. Oppenheimer, *Ann. Phys.* **1927**, *389*, 457–484.

[3] D. R. Hartree, *Math. Proc. Camb. Philos. Soc* **1928**, *24*, 89–110.

[4] J. C. Slater, *Phys. Rev.* **1930**, *35*, 210–211.

[5] V. Fock, *Z. Phys* **1930**, *61*, 126–148.

[6] D. Hait, M. Head-Gordon, *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.

[7] S. Kristyán, P. Pulay, *Chem. Phys. Lett.* **1994**, *229*, 175–180.

[8] J. Pérez-Jordá, A. Becke, *Chem. Phys. Lett.* **1995**, *233*, 134–137.

[9] O. Gunnarsson, B. I. Lundqvist, *Phys. Rev. B* **1976**, *13*, 4274–4298.

[10] D. Langreth, J. Perdew, *Solid State Commun.* **1975**, *17*, 1425–1429.

[11] D. C. Langreth, J. P. Perdew, *Phys. Rev. B* **1977**, *15*, 2884–2901.

[12] D. Bohm, D. Pines, *Phys. Rev.* **1953**, *92*, 609–625.

[13] H. Eshuis, J. Yarkony, F. Furche, *J. Chem. Phys.* **2010**, *132*, 234114.

[14] J. F. Dobson, J. Wang, B. P. Dinte, K. McLennan, H. M. Le, *Int. J. Quantum Chem.* **2005**, *101*, 579–598.

[15] J. F. Dobson, T. Gould, *J. Phys. Condens. Matter* **2012**, *24*, 073201.

[16] H. F. Schurkus, C. Ochsenfeld, *J. Chem. Phys.* **2016**, *144*, 031101.

[17] A. Luenser, H. F. Schurkus, C. Ochsenfeld, *J. Chem. Theory Comput.* **2017**, *13*, 1647–1655.

[18] A. M. Burow, J. E. Bates, F. Furche, H. Eshuis, *J. Chem. Theory Comput.* **2014**, *10*, 180–194.

[19] W. Klopper, A. M. Teale, S. Coriani, T. B. Pedersen, T. Helgaker, *Chem. Phys. Lett.* **2011**, *510*, 147–153.

[20] D. L. Freeman, *Phys. Rev. B* **1977**, *15*, 5512–5521.

[21] A. Grüneis, M. Marsman, J. Harl, L. Schimka, G. Kresse, *J. Chem. Phys.* **2009**, *131*, 154115.

[22] B. Mussard, D. Rocca, G. Jansen, J. G. Ángyán, *J. Chem. Theory Comput.* **2016**, *12*, 2191–2202.

[23] T. M. Henderson, G. E. Scuseria, *Mol. Phys.* **2010**, *108*, 2511–2517.

[24] R. Sedlak, T. Janowski, M. Pitonak, J. Rezac, P. Pulay, P. Hobza, *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374.

[25] W. Ritz, *J. Reine Angew. Math.* **1909**, *135*, 1–61.

[26] C. D. Sherrill, H. F. Schaefer, Advances in Quantum Chemistry **1999**, *34*, 143–269.

[27] C. C. J. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69–89.

[28] G. G. Hall, J. E. Lennard-Jones, *Proc. Royal Soc. A* **1951**, *205*, 541–552.

[29] S. F. Boys, A. C. Egerton, *Proc. Royal Soc. A* **1950**, *200*, 542–554.

[30] P.-O. Löwdin, *Phys. Rev.* **1955**, *97*, 1509–1520.

[31] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618–622.

[32] J. Čížek, *J. Chem. Phys.* **1966**, *45*, 4256–4266.

[33] J. Čížek, J. Paldus, L. Šroubková, *Int. J. Quantum Chem* **1969**, *3*, 149–167.

[34] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, *Chem. Phys. Lett.* **1989**, *157*, 479–483.

[35] M. W. Schmidt, M. S. Gordon, *Annu. Rev. Phys. Chem* **1998**, *49*, 233–266.

[36] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, K. Wolinski, *J. Phys. Chem.* **1990**, *94*, 5483–5488.

[37] K. Andersson, P. Malmqvist, B. O. Roos, *J. Chem. Phys.* **1992**, *96*, 1218–1226.

[38] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, *Chem. Phys. Lett.* **1998**, *286*, 243–252.

[39] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.

[40] R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, **1989**.

[41] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133–A1138.

[42] J. P. Perdew, K. Schmidt, *AIP Conf. Proc* **2001**, *577*, 1–20.

[43] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, M. Levy, *Phys. Rev. B* **1996**, *53*, 3764–3774.

[44] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 034108.

[45] E. Engel, R. M. Dreizler, *Density-Functional Theory, An Advanced Course*, Springer-Verlag, **2011**.

[46] R. P. Feynman, *Phys. Rev.* **1939**, *56*, 340–343.

[47] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200–1211.

[48] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[49] F. Furche, *J. Chem. Phys.* **2008**, *129*, 114105.

[50] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[51] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*, 9982–9985.

[52] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[53] J. Toulouse, F. Colonna, A. Savin, *Phys. Rev. A* **2004**, *70*, 062505.

[54] T. Leininger, H. Stoll, H.-J. Werner, A. Savin, *Chem. Phys. Lett.* **1997**, *275*, 151–160.

[55] J. Toulouse, F. Colonna, A. Savin, *J. Chem. Phys.* **2005**, *122*, 014110.

[56] T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett.* **2004**, *393*, 51–57.

[57] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, G. E. Scuseria, *J. Chem. Phys.* **2006**, *125*.

[58] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* **2003**, *118*, 8207–8215.

[59] J. Heyd, G. E. Scuseria, *J. Chem. Phys.* **2004**, *121*, 1187–1192.

[60] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* **2006**, *124*.

[61] A. Irmler, A. M. Burow, F. Pauly, *J. Chem. Theory Comput.* **2018**, *14*, 4567–4580.

[62] A. L. Fetter, J. D. Walecka, *Quantum Theory of Many-Particle Systems*, McGraw-Hill, **1971**.

[63] H. Lehmann, *Il Nuovo Cimento (1943-1954)* **1954**, *11*, 342–357.

[64] D. A. Strubbe, L. Lehtovaara, A. Rubio, M. A. L. Marques, S. G. Louie in *Fundamentals of Time-Dependent Density Functional Theory*, Springer Berlin Heidelberg, **2012**, pp. 139–166.

[65] E. K. U. Gross, W. Kohn, *Phys. Rev. Lett.* **1985**, *55*, 2850–2852.

[66] M. Gell-Mann, F. Low, *Phys. Rev.* **1951**, *84*, 350–354.

[67] G. C. Wick, *Phys. Rev.* **1950**, *80*, 268–272.

[68] R. P. Feynman, *Phys. Rev.* **1949**, *76*, 749–759.

[69] R. P. Feynman, *Phys. Rev.* **1949**, *76*, 769–789.

[70] A. Einstein, *Ann. Phys.* **1916**, *354*, 769–822.

[71] G. E. Scuseria, T. M. Henderson, D. C. Sorensen, *J. Chem. Phys.* **2008**, *129*, 231101.

[72] F. Furche, *Phys. Rev. B* **2001**, *64*, 195120.

[73] J. E. Bates, F. Furche, *J. Chem. Phys.* **2013**, *139*, 171103.

[74] J. Paier, X. Ren, P. Rinke, G. E. Scuseria, A. Grüneis, G. Kresse, M. Scheffler, *New J. Phys.* **2012**, *14*, 043002.

[75] X. Ren, P. Rinke, G. E. Scuseria, M. Scheffler, *Phys. Rev. B* **2013**, *88*, 035120.

[76] E. E. Salpeter, H. A. Bethe, *Phys. Rev.* **1951**, *84*, 1232–1242.

[77] A. Heßelmann, A. Görling, *Mol. Phys.* **2010**, *108*, 359–372.

[78] P. Bleiziffer, D. Schmidtel, A. Görling, *J. Chem. Phys.* **2014**, *141*, 204107.

[79] P. Pulay, *Chem. Phys. Lett.* **1983**, *100*, 151–154.

[80] M. Kállay, *J. Chem. Phys.* **2015**, *142*, 204105.

[81] A. Heßelmann, *J. Chem. Phys.* **2017**, *146*, 174110.

[82] M. Feyereisen, G. Fitzgerald, A. Komornicki, *Chem. Phys. Lett.* **1993**, *208*, 359–363.

[83] J. L. Whitten, *J. Chem. Phys.* **1973**, *58*, 4496–4501.

[84] E. Baerends, D. Ellis, P. Ros, *Chemical Physics* **1973**, *2*, 41–51.

[85] C. Van Alsenoy, *J. Comput. Chem.* **1988**, *9*, 620–626.

[86] H.-J. Werner, F. R. Manby, P. J. Knowles, *J. Chem. Phys.* **2003**, *118*, 8149–8160.

[87] H. F. Schurkus, A. Luenser, C. Ochsenfeld, *J. Chem. Phys.* **2017**, *146*, 211106.

[88] M. Häser, R. Ahlrichs, *J. Comput. Chem.* **1989**, *10*, 104–111.

[89] J. D. Cloizeaux, *Phys. Rev.* **1964**, *135*, A685–A697.

[90] J. J. Rehr, W. Kohn, *Phys. Rev. B* **1974**, *10*, 448–455.

[91] W. Kohn, J. R. Onffroy, *Phys. Rev. B* **1973**, *8*, 2485–2495.

[92] C. Ochsenfeld, C. A. White, M. Head-Gordon, *J. Chem. Phys.* **1998**, *109*, 1663–1669.

[93] C. Ochsenfeld, *Chem. Phys. Lett.* **2000**, *327*, 216–223.

[94] J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2013**, *138*, 134114.

[95] J. Almöf, *Chem. Phys. Lett.* **1991**, *181*, 319–320.

[96] M. Häser, J. Almlöf, *J. Chem. Phys.* **1992**, *96*, 489–494.

[97] M. Häser, *Theor. Chem. Acc* **1993**, *87*, 147–173.

[98] S. A. Maurer, L. Clin, C. Ochsenfeld, *J. Chem. Phys.* **2014**, *140*, 224112.

[99] S. Vogler, M. Ludwig, M. Maurer, C. Ochsenfeld, *J. Chem. Phys.* **2017**, *147*, 024101.

[100] P. Y. Ayala, G. E. Scuseria, *J. Chem. Phys.* **1999**, *110*, 3660–3671.

[101] S. Schweizer, B. Doser, C. Ochsenfeld, *J. Chem. Phys.* **2008**, *128*, 154101.

[102] B. Doser, D. S. Lambrecht, C. Ochsenfeld, *Phys. Chem. Chem. Phys.* **2008**, *10*, 3335–3344.

[103] M. Kaltak, J. Klimeš, G. Kresse, *J. Chem. Theory Comput.* **2014**, *10*, 2498–2507.

[104] H. N. Rojas, R. W. Godby, R. J. Needs, *Phys. Rev. Lett.* **1995**, *74*, 1827–1830.

[105] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. White, R. Godby, *Comput. Phys. Commun* **2000**, *125*, 105–118.

[106] N. J. Higham, *WIREs Comp Stats* **2009**, *1*, 251–254.

[107] H. Harbrecht, M. Peters, R. Schneider, *Appl Numer Math* **2012**, *62*, 428–440.

[108] J. Zienau, L. Clin, B. Doser, C. Ochsenfeld, *J. Chem. Phys.* **2009**, *130*, 204112.

[109] H. Koch, A. S. de Merás, T. B. Pedersen, *J. Chem. Phys.* **2003**, *118*, 9481–9484.

[110] A. Luenser, J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2016**, *145*, 124103.

[111] F. Aquilante, T. Bondo Pedersen, A. Sánchez de Merás, H. Koch, *J. Chem. Phys.* **2006**, *125*, 174101.

[112] J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2007**, *127*, 054103.

[113] F. G. Gustavson, *ACM Trans. Math. Softw.* **1978**, *4*, 250–269.

[114] P. R. Surján, *Chem. Phys. Lett.* **2005**, *406*, 318–320.

[115] B. Ramberger, T. Schäfer, G. Kresse, *Phys. Rev. Lett.* **2017**, *118*, 106403.